

Chapter 6

Data-Mining Techniques for an Analysis of Non- Conventional Methodologies: Deciphering of Alternative Medicine

William Claster

Ritsumeikan Asia Pacific University, Japan

Nader Ghotbi

Ritsumeikan Asia Pacific University, Japan

Subana Shanmuganathan

Auckland University of Technology, New Zealand

ABSTRACT

Some common methodologies in our everyday life are not based on modern scientific knowledge but rather a set of experiences that have established themselves through years of practice. As a good example, there are many forms of alternative medicine, quite popular, however difficult to comprehend by conventional western medicine. The diagnostic and therapeutic methodologies are very different and sometimes unique, compared to that of western medicine. How can we verify and analyze such methodologies through modern scientific methods? We present a case study where data-mining was able to fill this gap and provide us with many tools for investigation. Osteopathy is a popular alternative medicine methodology to treat musculoskeletal complaints in Japan. Using data-mining methodologies, we could overcome some of the analytical problems in an investigation. We studied diagnostic records from a very popular osteopathy clinic in Osaka, Japan that included over 30,000 patient visits over 6 years of practice. The data consists of some careful measurements of tissue electro-conductivity differences at 5 anatomical positions. Data mining and knowledge discovery algorithms were applied to search for meaningful associations within the patient data elements recorded. This study helped us scientifically investigate the diagnostic methodology adopted by the osteopath.

DOI: 10.4018/978-1-60566-266-4.ch006

INTRODUCTION

Growing acceptance of alternative medicine by the public has convinced third party organizations to increase the level of insurance coverage and has caused the U.S. Congress to rule for establishment of the National Center for Complementary and Alternative Medicine (NCCAM) at the National Institutes of Health (NIH) (Jonas, 1998; Nahin & Straus, 2001). However, these methodologies are still considered largely unsubstantiated by many practitioners of conventional western medicine. The problem many of these investigators are faced with is the difficulty in performing trials of an intervention not based on clinical practice as well as trials of multifaceted interventions that are too complicated for a conventional study design (Miller et al, 2004; Lewith, 2003). Can data-mining help explore non-conventional methodologies such as those used in alternative medicine in the light of science? NCCAM has been seeking the advice of the scientific research community in this regard.

Western medicine has benefited from inquiry into non-traditional healing practices whether it has been through the study of chemical agents or through the study of physical modalities like acupuncture. Complimentary/alternative medicine (CAM) is gaining more popularity in the US, Europe, Australia and elsewhere (Eisenberg et al, 1998; Cassileth et al, 2001). Probably the popularity of complementary /alternative medicine has increased because of some dissatisfaction with modern medicine or with the economics of clinical care surrounding the management of ill-defined chronic conditions (Imanishi et al, 1999). The following classification has been suggested for the main CAM modalities (Gordon et al, 1998): manual therapies (e.g. chiropractic, osteopathy, acupuncture, acupressure, and massage), oral therapies (e.g. herbal medicines, homeopathy, diets, and vitamins), mind-body therapies (e.g. meditation, relaxation, biofeedback, and hypnosis), movement-based therapies (e.g. Tai chi),

and support therapies (e.g. counseling, support groups, prayer, and other religious practices). Manual therapies are by far the most widely used CAM modalities (Koes et al, 1992; Eisenberg et al, 1993; MacLennan et al, 1996; Astin, 1998; Druss & Rosenheck, 1999). Although there are many issues that arise in categorizing the use of diagnostic modalities as western versus non-western, we will not concern ourselves with such classification issues in this paper any further

BACKGROUND

In Japan, Chinese herbal medicine (kampo) which was originally introduced in the 5th and 6th century has been significantly modified by Japanese practitioners over a long time. Kampo was excluded from authorized medical practice about 100 years ago but is still, along with acupuncture, electroacupuncture and moxibustion widely practiced and popular. In Japan, non-western medical treatments are sought for various ailments and illnesses in spite of the fact that medical practice in Japan is one of the most advanced in the world.

Modern osteopathy was probably started in the late 1800s by an American physician called Andrew Still. Osteopathy is now a quite established profession in the U.S.A. with many osteopaths diagnosing and treating medical problems using manual touch. They get training on how to feel (through palpation) the body's anatomy, the texture and motion of tissues, the flow of fluids and its structural makeup. In osteopathy, the body's innate power to heal itself is emphasized, and it is believed that previous physical trauma leaves its touch on the body's structure. The osteopaths generally try to develop a strong sense of touch to detect physical problems, and to apply the exactly right amount of pressure to treat dysfunction in the motion of the tissues, restore movement of fluids and to release compressed joints and bones.

In Japan, a particular type of osteopathy that has its roots in the martial arts practice of Judo

is also widespread and in contrast to most other non-conventional medical practices has achieved acceptance into the medical health system. In fact, some alternative practices in this area are recognized by the national health program and are covered by insurance plans. There are many schools of thought in osteopathy and some practitioners develop their own unique way of palpation and detection of anatomical problems.

We were informed of a very popular osteopathy practice in Osaka, Japan which also kept electronic records of the patients on their computers. Therefore we contacted the main practitioner and got an opportunity to visit the practice, hear about the methodology of examination and diagnosis, and also collect the large amount of recorded data. We also explained our data mining methodology and the fact that we were interested in a non-biased study of the data and any patterns of associations among them which could help decipher the methodology and make it more understandable to conventional care. The sorts of data that were available and their way of interpretation will be explained in the following sections.

The data gathered consisted of electro-conductivity measurements taken at five anatomical positions which include the neck, armpits, wrists, knees, and ankles. These recordings were used as a diagnostic tool by the osteopath practitioner to isolate the primary problem area of the patient. The practitioner would take readings from the right and left side of the patient at each of these five locations and would consider the measures of difference between the left and right at all locations in one series of recordings. The location which showed the greatest imbalance would be deemed as the initial problem location. From that location, the practitioner would approach to introduce "energy" and physical manipulations to help the patient.

Although our efforts fell short of clinical verification, we were able to identify useful patterns using self organizing maps, decision trees, and association and sequence rule mining approaches.

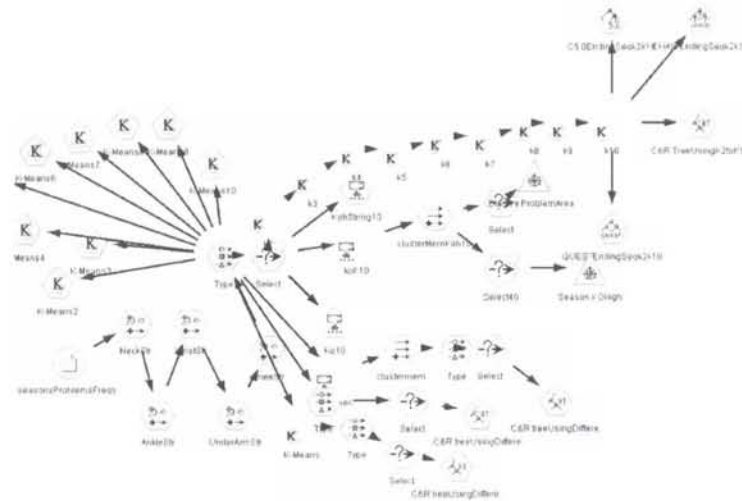
We also studied the attributes of a considerable number of users of osteopathy services which can be useful in the delineation of the social position of this alternative medicine modality in Japan.

DATA MINING METHODOLOGY

We explored the data using self-organizing maps (SOM) and web graphs. A self-organizing map (SOM) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map. The map seeks to preserve the topological properties of the input space. This makes SOM useful for visualizing low-dimensional views of high-dimensional data, akin to multidimensional scaling. The model was first described as an artificial neural network by Teuvo Kohonen, and is sometimes called a Kohonen map. We also explored the sequential patterns in the data using a sequential form of an association algorithm. Together with the practitioner we defined a 'balanced' state. We were able to find sequences of readings with a support of greater than 30% and confidence level of 70%. These association rules identified a few likely sequences of readings that lead to this balanced state. We then built a predictive decision tree using the C5 algorithm which included inputs from the association rules, the original data, and the clusters obtained from the SOM maps.

We used Self-organizing maps to analyze the electro-conductivity readings. The practitioner provided us with approximately 32,077 patient visit records. At each visit the practitioner used an electronic measuring machine which was purchased locally in Japan and modified to enable convenient measurements in the doctor-patient setting. The readings consisted of both a resistance measurement (RES) and a pulsed direct current (PDC) measurement for unidirectional currents, taken at the left and right sides of each of the five anatomical locations on the patient. The range

Figure 1. The Clementine workspace: Each node represents a step in the analysis. In our analysis we used both K-means and Kohonen clustering algorithms. Kohonen maps and K-means are often considered to be equally stable and accurate



of electro-conductive difference at each of these locations was 8.31, the mean was 0.07, and the standard deviation was 0.99. The readings were all distributed in an almost normal distribution with the largest skew coefficient being -0.19

Based on the suggestion of the practitioner we used only the electro-conductive measurements for our initial investigation. The data was then normalized and categorized based on measurements taken from a normal distribution and using the mean and standard deviation from each location. This produced measurements for each location that were categorized within the set $\{-3, -2, -1, 0, 1, 2, 3\}$ where -3 indicated the most extreme imbalance towards the left and 3 indicating the most extreme imbalance to the right. We tried various Kohonen SOM maps and measured their silhouette coefficients.

In this Chapter we present the results of analytical investigation using Clementine data mining package. This package proved to be a convenient tool for step by step exploration of the data. It provides a visual map of the process of exploration. Figure 1 shows a visual map of the steps in

our investigation. Each 'node' represents either a step in the process or else a particular visualization of the data.

Kohonen Self Organizing Maps (SOM) and Clementine Web Graphs

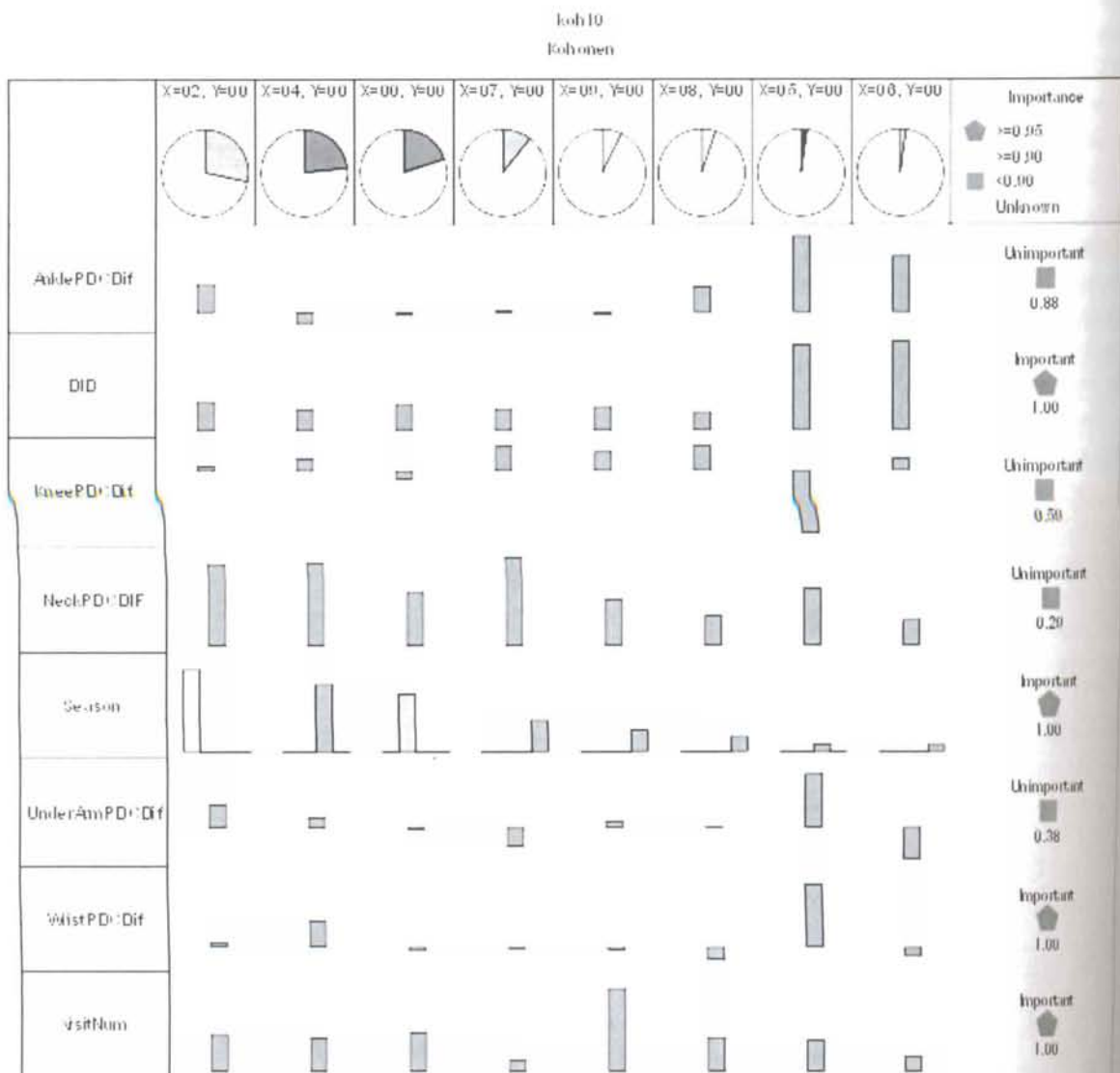
We explored the Kohonen clusters from 2 to 15 clusterings (Figure 2). We then used the silhouette coefficient (Knuth, 1973). The silhouette coefficient is a measure of the clustering quality that is independent from the number of clusters (Kaufman & Rousseeuw, 1990). It measures the cohesiveness of the items within a group as compared to the dissimilarity of the points within different clusters. The silhouette coefficient was largest for 8 clusters at 0.72. This indicates a clustering result with very good separation among clusters, viz. Data points are very close to the center of their cluster and remote from the next nearest cluster. We also tried using the K-means algorithm. The results differed from Kohonen but the silhouette coefficients did not vary as much and the highest silhouette coefficient was 0.4 with 5 clusters.

We identified certain clusters as having distinct characteristics. In particular two clusters were identified as having a large number of summer visits. In one of those clusters the problem areas were located in one group of locations (wrists,

underarm, and knee) whereas the other cluster had a different set of problem areas (neck and underarm).

Clementine has a visualization technique referred to as a web graph. It visualizes associations

Figure 2. Kohonen SOM for 8 clusters: In Clementine clusters are named with their x and y coordinates of the associated nodes.. Cluster x=04, y=00, and cluster x=05,y=00 were both identified clusters of patient visits in the summer but differed in other characteristics. Web graphs revealed relationships within the clusters. Cluster x=05,y=00 had patients with a problem area primarily located in the wrist, underarm, or knee, whereas cluster x=04, y=00 had patient visits where the problem area was measured as being located primarily in the neck and underarm (Source: Authors)



between different values of a collection of (at least 2) nominal variables. Darker lines in the web graph indicate stronger relations between outcomes that are connected by the line. For example, as seen in the graph in Figure 3 there are dark lines connecting a value of the season variable (summer) with 3 separate outcomes a variable indicating the problem area (knee, neck, and underarm). This is because this was a web graph of cluster $x=04$, $y=00$, whereas Figure 4 is a web graph of cluster $x=05$, $y=00$. In cluster $x=05$, $y=00$ summer was associated with underarm and neck problem areas.

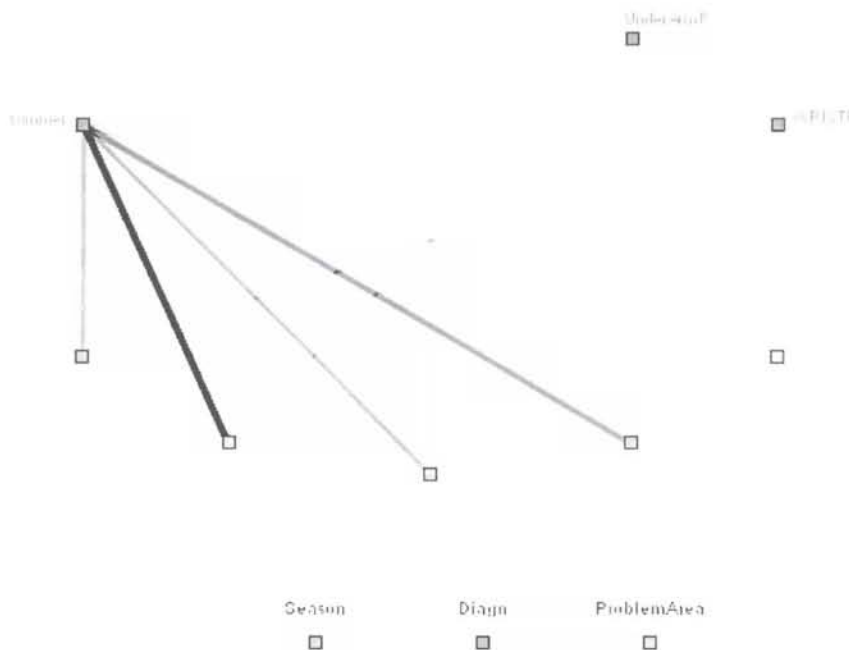
Underarm was overall the most prevalent problem area measured and so the practitioner will focus on the neck versus the wrist in these two clusters. The practitioner wanted to use this to distinguish between two possible types of patient visits and look for variations between responses to treatments in the summer between patients who have neck located problems versus

knee or wrist problems. Thus the clustering and further exploration of different groups discovered through clustering may lead to focused and more effective treatment.

MODELING THE DATA TO PREDICT A BALANCED STATE

We used C5.0 algorithm to predict whether the patient's next visit would result in a balanced state (see the decision tree presented in Figure 5 below). A balanced state was defined to be a visit where the maximum imbalance across all the readings was within 0.25 standard deviations of the mean. The inputs to the model are Problem Area, visit number, and season. The output is the yes/no variable indicating whether a balanced state had been achieved. The correlation between a balance state, thus defined, and the lack of symptoms

Figure 3 Clementine web graph: This is a web graph of the data from cluster three variables (season, diagnosis, and problem area) are pictured and their possible values. Diagnosis was the original location of the problem area as measured by PDC and problem area was the location at the current visit. We see the boldest lines connecting summer and underarm, summer and neck, and summer and knee



was only 61%. It is hoped that this model may help to better understand patients who reached a balanced state and yet still had symptoms. C5.0 produces a decision tree which can provide an intuitive model for predictions. Consultation with the practitioner lead us to believe that this would be the best choice for reasons pertaining to its usability by the practitioner since unlike neural networks which are presented as “black box”, decision trees present the practitioner with an intuitive feel for the model. The predictions for the outcome “No” (meaning ‘not balanced’ as defined above) did not come with as high an average confidence level as those of ‘Yes’. The confidence for ‘Yes’ was above 75%.

The output variable is a dichotomous nominal variable, whether the measurement is characterized as balanced (Yes) or not balanced (No). The

pair of numbers in brackets to the right of each branch of the tree indicates (number of records, confidence). From the C5.0 decision tree we can discover various rules. For example, the model predicts that if the problem area is in the knee then there is 75 percent chance that the next visit will result in a balanced reading; on the other hand, if the problem area is in the neck, the season is spring, and the number of visits to the practitioner is greater than 32, then the next visit will result in a balanced reading with 100 percent confidence; again on the other hand, if the problem area is neck, the season winter, and the number of visits less than 5, then there is a 75 percent chance of a balanced state, etc. These rules should help the practitioner to get more insight into the prognosis and treatment of their patients.

Figure 4. Clementine web graph: This is a web graph of cluster x=04, y=00. This cluster had patient visits where the problem area was measured as being located primarily in the neck and underarm

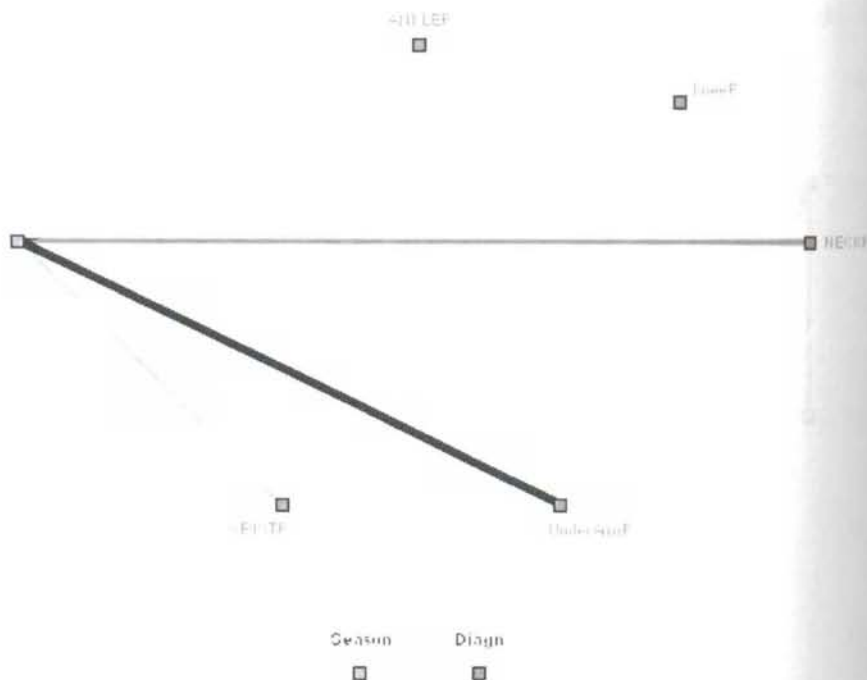


Figure 5. C5.0 decision tree

```

ProblemArea = ANKLEP [Mode: No] → No
- ProblemArea = KneeP [Mode: Yes]
  visitNum <= 1 [Mode: No] → No
  visitNum > 1 [Mode: Yes] → Yes
- ProblemArea = NECKP [Mode: No]
  Season = fall [Mode: No] → No
  - Season = spring [Mode: No]
    visitNum <= 32 [Mode: No] → No
    visitNum > 32 [Mode: Yes] → Yes
  Season = summer [Mode: No] → No
  - Season = winter [Mode: No]
    - visitNum <= 118 [Mode: No]
      visitNum <= 5 [Mode: Yes] → Yes
      visitNum > 5 [Mode: No] → No
    visitNum > 118 [Mode: Yes] → Yes
- ProblemArea = UnderArmP [Mode: Yes]
  Season = fall [Mode: No] → No
  Season = spring [Mode: No] → No
  Season = summer [Mode: No] → No
  Season = winter [Mode: Yes] → Yes
ProblemArea = WRISTP [Mode: No] → No

```

CONCLUSION

Alternative medicine includes a variety of different practices; however, they share a common feature in that modern clinical methods cannot be used to properly comprehend them without causing a lot of controversy and disagreement on both sides. Although it has been suggested that modern study methods can investigate but the experts of the alternative medical practices may disagree with this approach. Even the definitions of health vary. However using data mining techniques we endeavor to analyze the methodology without interfering with the practitioner. Meanwhile, many of these alternative medicine modalities are built upon overwhelming set of data in respect with their breadth and detail. We offered an example from a popular osteopathy clinic in Osaka which contained tens of thousand of measured data. Clustering of these alternative patient records helped us to see groups of similar patients or symptoms. In our study we found two groups of patients who exhibited similar symptoms

in certain seasons. On one hand, we were able to examine this unorthodox methodology based on a series of data mining methods including Kohonen networking, decision analysis trees, etc.

On the other hand, the alternative medical practitioner can use the gained knowledge to better understand the associations discovered amongst patients. They can develop modifications in treatment methods for different groups in order to see if they respond uniformly within the groups. Similar explorations can be made for other seasons or across other variables.

Building a model for alternative healthcare is a daunting task because of the novelty and complexity of the data and because the stakes are so high. On the other hand, because the stakes are so high there is strong impetus to formulate such models. In our work, there was an inherent limitation on our access to clinically comparable data such as the medical complaints of the patients and clear indications of a response to modern treatments modalities. However, we were able to examine the possibility of building a model based on the

concepts of the osteopath practitioner himself in respect with balance and imbalance. With this improvised definition of balance we were able to build a model which could help the practitioner to predict a prognosis of the problem and whether the next state of the patient would be more balanced based on the current state.

The approach of this osteopathy methodology is somewhat unique in attempting to blend the intuitive 'energy' approach of complimentary, traditional medicine with scientific record keeping and measurements. We hope to continue a discussion of the findings with the practitioners of this osteopathy clinic and to encourage a measurement across more variables like initial patient complaint, age, gender, and other physical bodily measurements. With additional data such as these, the model building can be more significant.

REFERENCES

- Astin, J. A. (1998). Why patients use alternative medicine: results of a national study. *Journal of the American Medical Association*, 279, 1548-1553. doi:10.1001/jama.279.19.1548
- Cassileth, B. R., Schraub, S., Robinson, E., & Vickers, A. (2001). Alternative medicine use worldwide: the International Union against Cancer survey. *Cancer*, 91(7), 1390-1393. doi:10.1002/1097-0142(20010401)91:7<1390::AID-CNCR1143>3.0.CO;2-C
- Druss, B. G., & Rosenheck, R. A. (1999). Association between use of unconventional therapies and conventional medical services. *Journal of the American Medical Association*, 282, 651-656. doi:10.1001/jama.282.7.651
- Eisenberg, D. M., Davis, R. B., Ettner, S. L., Appel, S., Wilkey, S., Van Rompay, M., & Kessler, R. C. (1998). Trends in Alternative Medicine Use in the United States, 1990-1997: Results of a Follow-up National Survey. *Journal of the American Medical Association*, 280, 1569-1575. doi:10.1001/jama.280.18.1569
- Eisenberg, D. M., Kessler, R. C., Foster, C., Norlock, F. E., Calkins, D. R., & Delbanco, T. L. (1993). Unconventional medicine in the United States: Prevalence, costs, and patterns of use. *The New England Journal of Medicine*, 328, 246-252. doi:10.1056/NEJM199301283280406
- Gordon, N. P., Sobel, D. S., & Tarazona, E. Z. (1998). Use of and interest in alternative therapies among adult primary care clinicians and adult members in a large health maintenance organization. *The Western Journal of Medicine*, 169, 153-161.
- Imanishi, J., Watanabe, S., Satoh, M., & Ozasa, K. (1999). Japanese doctors' attitudes to complementary medicine. *Lancet*, 354(9191), 1735-1736. doi:10.1016/S0140-6736(05)76729-5
- Jonas, W. (1998). Alternative Medicine and the Conventional Practitioner. *Journal of the American Medical Association*, 279, 708-709. doi:10.1001/jama.279.9.708
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data*. New York: John Wiley & Sons, Inc.
- Knuth, D. E. (1973). *The Art of Computer Programming: Sorting and Searching*, 3. MASS: Addison-Wesley.
- Koes, B. W., Bouter, L. M., Van Mameren, H., Es-sers, A. H., Verstegen, G. M., & Hofhuizen, D. M. (1992). Randomized clinical trial of manipulative therapy and physiotherapy for persistent back and neck complaints: results of 1 year follow up. *British Medical Journal*, 304, 601-605. doi:10.1136/bmj.304.6827.601
- Lewith, G. (2003). Can we evaluate electrodermal testing? *Complementary Therapies in Medicine*, 11(2), 115-117. doi:10.1016/S0965-2299(03)00057-8

MacLennan, A. H., Wilson, D. H., & Taylor, A. W. (1996). Prevalence and cost of alternative medicine in Australia. *Lancet*, *347*, 569–573. doi:10.1016/S0140-6736(96)91271-4

Miller, F. G., Emanuel, E. J., Rosenstein, D. L., & Straus, S. E. (2004). Ethical Issues Concerning Research in Complementary and Alternative Medicine. *Journal of the American Medical Association*, *291*, 599–604. doi:10.1001/jama.291.5.599

Nahin, R. L., & Straus, S. E. (2001). Research into complementary and alternative medicine: problems and potential. *British Medical Journal*, *322*, 61–164. doi:10.1136/bmj.322.7279.161