original research

Advanced Treatment Monitoring for Olympic-Level Athletes Using Unsupervised Modeling Techniques

Jacob A. Siedlik, MA*; Charles Bergeron, PhD†; Michael Cooper, MS*; Russell Emmons, MS*; William Moreau, DC, DACBSP‡; Dustin Nabhan, DC, DACBSP, CSCS‡; Philip Gallagher, PhD*; John P. Vardiman, PhD, LAT, ATC*||

*Department of Health, Sport, and Exercise Sciences, University of Kansas, Lawrence; †Albany College of Pharmacy and Health Sciences, NY; ‡US Olympic Committee, Colorado Springs; ||Department of Human Nutrition, Kansas State University, Manhattan

Context: Analysis of injury and illness data collected at large international competitions provides the US Olympic Committee and the national governing bodies for each sport with information to best prepare for future competitions. Research in which authors have evaluated medical contacts to provide the expected level of medical care and sports medicine services at international competitions is limited.

Objective: To analyze the medical-contact data for athletes, staff, and coaches who participated in the 2011 Pan American Games in Guadalajara, Mexico, using unsupervised modeling techniques to identify underlying treatment patterns.

Design: Descriptive epidemiology study.

Setting: Pan American Games.

Patients or Other Participants: A total of 618 US athletes (337 males, 281 females) participated in the 2011 Pan American Games.

Main Outcome Measure(s): Medical data were recorded from the injury-evaluation and injury-treatment forms used by clinicians assigned to the central US Olympic Committee Sport Medicine Clinic and satellite locations during the operational 17-day period of the 2011 Pan American Games. We used principal components analysis and agglomerative clustering algorithms to identify and define grouped modalities. Lift statistics were calculated for within-cluster subgroups.

Results: Principal component analyses identified 3 components, accounting for 72.3% of the variability in datasets. Plots of the principal components showed that individual contacts focused on 4 treatment clusters: massage, paired manipulation and mobilization, soft tissue therapy, and general medical.

Conclusions: Unsupervised modeling techniques were useful for visualizing complex treatment data and provided insights for improved treatment modeling in athletes. Given its ability to detect clinically relevant treatment pairings in large datasets, unsupervised modeling should be considered a feasible option for future analyses of medical-contact data from international competitions.

Key Words: statistical methods, epidemiology, treatment patterns, clinician care decisions

---

Key Points
- Unsupervised modeling techniques were useful for identifying treatment modalities that share multiple attributes.
- Unsupervised modeling provided a macroperspective of clinically relevant treatment pairings and a guide for future resource allocation.

Injury and illness surveillance continues to play a pivotal role in maintaining athlete health and providing individualized injury-prevention programs.[1–3] The systematic monitoring of injuries and illnesses occurring and treatment modalities used at large-scale international competitions can provide insight into how to allocate resources and recruit personnel for future events.[4] In addition, medical representatives serving on the local organizing committee (LOC), established to act as the liaison between the games and the host city, can use these data to ensure that the medical community and hospitals of the host city are prepared for the infusion of athletes and staff[5–7] by having the appropriate equipment, facilities, and staff for necessary health care procedures.[8]

Traditional analysis for these events focuses on determining the rates of injuries and illness that occur during competition.[4] Researchers[4,9–13] have documented injuries and illness for athletes participating at numerous international competitions, specifically focusing on overall surveillance. These data are often collected from the medical staff of a national team of a participating country and from the polyclinic of the host.[4] They do not provide a detailed description of the medical care provision for each injury or illness. Furthermore, they do not account for the preventive and recovery treatments provided by the medical staff of the national team. No investigators have documented the medical treatments provided during patient contacts or conducted analyses of treatment patterns and methods used in the high-volume clinics at international competitions. Therefore, the purpose of our study was to catalog and analyze 1957 medical contacts between US team members and clinicians at the Pan American Games. Our primary purpose was to show how dimensionality-reduction techniques can be applied to clinical data to

Table 1. Available Treatment Modalities at the 2011 Pan American Games

Treatment Modalities
  Bracing
  Computed tomography scan
  Diagnostic ultrasound
  Electrical muscular stimulation
  Graston technique
  Ice
  Laboratory tests
  Magnetic resonance imaging
  Manipulation
  Massage
  Mobilization
  Moist hot pack
  NormaTec[a]
  Other
  Proprioceptive neuromuscular facilitation
  Radiograph
  Soft tissue manipulation
  Stretch
  Taping
  Ultrasound

[a] NormaTec, NormaTec Recovery Systems, Newton Center, MA.

reveal trends that are difficult to visualize using standard descriptive statistics. We hypothesized that, by using unsupervised modeling methods, we would identify clinically relevant treatment patterns that aid in developing strategies for resource allocation, personnel assignments, and clinical skills required at large-scale athletic competitions.

METHODS

Participants

To address the lack of information related to treatment patterns, we collected the medical-contact data recorded at the US Olympic Committee (USOC) Sports Medicine Clinic during the 2011 Pan American Games. More than 600 US athletes, coaches, and staff participated in these games. The multivariate nature of the data acquired from the clinic (ie, many attributes recorded for each medical contact) leads to difficulty in visualizing patterns and is a problem characteristic of large datasets in general. This sample represents the number of US personnel that travel to large-scale competitions; we chose it specifically to test the feasibility of concept for using unsupervised-modeling techniques to examine treatment data and uncover clinically relevant attributes. A total of 618 US athletes (337 males, 281 females) participated in the 2011 Pan American Games. The operational timeframe spanned the opening of the games on October 14, 2011, through the closing on October 30, 2011. During this 17-day period, 77 practitioners representing both traditional and alternative medical disciplines provided care at 35 sport venues across Guadalajara and Puerto Vallarta as part of the USOC Sports Medicine Clinic.

Data Collection

The clinicians recorded medical-contact information for all injuries and illnesses on the Injury Evaluation and Treatment Form used by the USOC. After competition, records of medical contacts were transposed and coded for analysis. Descriptive information about the injuries or illnesses was based on the chief concern recorded on the Injury Evaluation and Treatment Form. Data were recorded for acute and chronic conditions. After the records were transcribed, the data were blinded for analysis. Either no contacts were available or no data were provided for sports including fencing, rifle, rugby, softball, trap/skeet, women's water polo, and women's volleyball. The study was approved by the Human Subjects Committee of the University of Kansas at Lawrence.

Statistical Analysis

Data were analyzed using MATLAB (version 8.1 R2013a; The MathWorks Inc, Natick, MA). Any combination of the 20 possible treatments existed for each contact (Table 1). These were stored as a 1957×20 data matrix, with each row representing a contact between a team member and a clinician and each column representing a treatment. The matrix entry for row i and column j was 1 if contact i involved treatment j and 0 if otherwise. Data for diagnostic testing (eg, radiograph, magnetic resonance imaging) were recorded but are not presented in this analysis. We performed principal components analysis (PCA) on these data. Initially, clusters were identified manually using separating planes, permitting easy assignment of contacts to clusters. Afterward, we realized that agglomerative clustering finds equivalent rules with less user intervention and assigns a few outlier contacts to singleton clusters.

Clusters were interpreted as treatment modalities and further described in 2 ways. First, the treatment profile associated with the mean point of each cluster could be mapped back to treatment space, so we knew how treatments associated with clusters. Second, relationships were found between clusters and contact metadata: sex, position, sport, condition, and provider type. The lift was calculated for each cluster and metadata label, and 1-sided exact tests for goodness of fit[14] at an α level of .05 determined which labels were overrepresented and underrepresented in each cluster.

RESULTS

The mean number of treatments per contact was 1.34, with a mode and median of 1. The most common treatments were massage (n = 888, 45.4% of encounters), soft tissue manipulation (STM; n = 535, 27.3%), chiropractic manipulation (CM; n = 293, 15.0%), and joint mobilization (JM; n = 222, 11.3%). By applying PCA, we found the principal components (PCs); the relative variances are provided in Table 2. The leading 3 PCs are plotted in Figure 1. Taken together, these account for 72.3% of the data variance. Each point in Figure 1 corresponds to a patient-clinician contact. In this view, a clear trend emerges in the form of 4 clusters.

The 4 clusters are defined fully in Table 3 and summarized in Table 4. In Table 3, we report the lift for each cluster and metadata label pair. For clarity, we provide a brief example. The empirical likelihoods of a contact belonging to each cluster were 0.454, 0.121, 0.177, and 0.248, respectively. These numbers are calculated by dividing the observed cluster size by the total number of

Table 2. Relative Variability for the 20 Components in the Data Matrix

Component	Relative Variance, %
1	33.85
2	29.90
3	8.53
4	6.17
5	4.49
6	4.13
7	3.52
8	1.89
9	1.72
10	1.51
11	1.27
12–20	Each <1

contacts in the dataset (ie, cluster I: 888/1957 = 0.454). The listed likelihoods intuitively make sense when one considers how they are calculated. Cluster I had 888 contacts and represented 45.4% of the overall dataset; therefore, we would expect approximately 45% of the shoulder injuries to fall into this cluster. In actuality, however, the 114 contacts reporting a shoulder/upper arm injury were distributed across the 4 clusters as follows: 1/114 = 0.00877, 17/114 = 0.149, 57/114 = 0.5, and 39/114 = 0.342, or 0.87%, 14.9%, 50%, and 34.2% of the injuries within the respective clusters. The shoulder/upper arm injury lift statistics for each cluster are calculated by dividing the observed distribution of shoulder/upper arm injuries appearing within a cluster by the expected distribution. The calculations for each cluster are as follows: 0.00877/0.454 = 0.02, 0.149/0.121 = 1.23, 0.5/0.177 = 2.82, and 0.342/0.248 = 1.38. Using the left-sided exact test for goodness of fit,[14] we observed that 1 contact

of 114 rejected the null hypothesis that the proportion of cluster I contacts was 0.454 or more against or in favor of the alternative hypothesis that the proportion was less than 0.454 at an α level of .05. We interpret this to mean that shoulder/upper arm injuries were underrepresented in cluster I. This underrepresentation does not mean the injuries were not being reported; instead it shows that the injuries were not distributed randomly across the clusters and provides evidence for an underlying structure in the observed data. This underrepresentation is provided in Table 3. Similarly, cluster and metadata label pairs that rejected the right-sided null hypothesis are noted in Table 3, meaning that a metadata label is overrepresented in a cluster. Similar to the interpretation of underrepresentation, labels that are overrepresented indicate that they appear at higher values in a given cluster than we would expect given a random distribution. We now discuss each cluster.

Cluster I: Massage

Cluster I was the most homogeneous cluster. It showed that 888 contacts involved a massage, and an overwhelming majority (96.5%) were limited to that treatment. The presence or absence of a massage also provided a meaning to PC 1 score; that is, a value of approximately 1 identified contact, including a massage, and a value of approximately 0 did not. Cluster I consisted of a tight core of 857 massage-only contacts; a smaller number (31 contacts) involving massage and an additional treatment were scattered about the core of cluster I (Figure 1). The mean cluster I contact involved massage; all other treatments presented less than 2% incidence each. This cluster was overrepresented by cyclists and swimmers. Massage therapists were the predominant clinicians (646 contacts), and virtually all
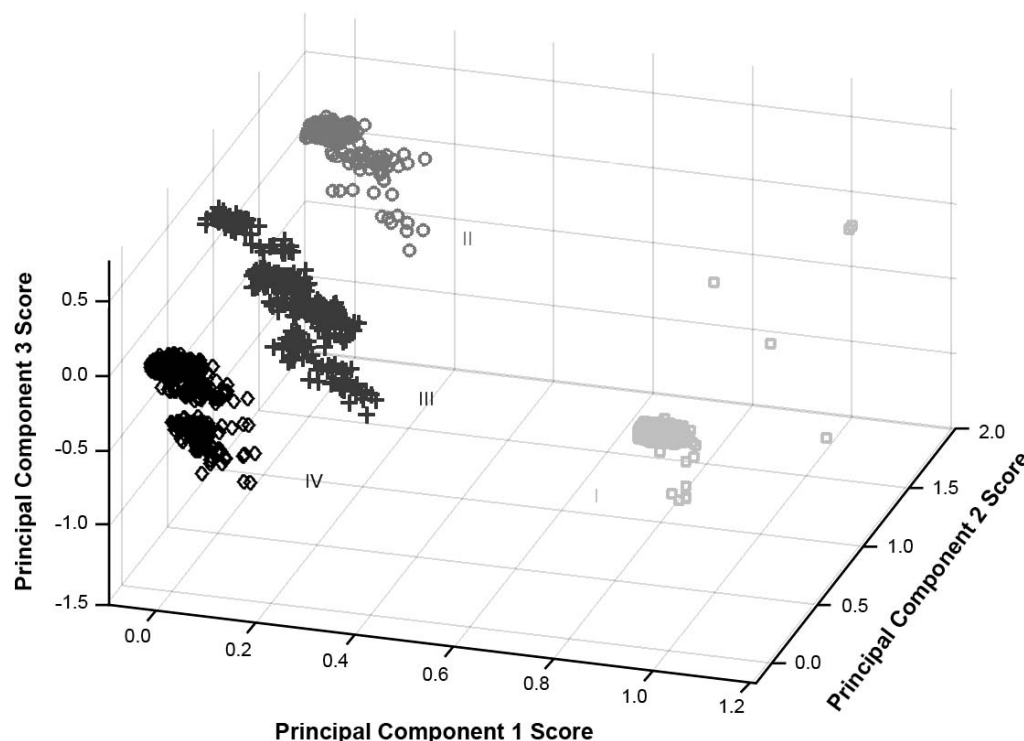

Figure 1. Principal component analysis tri-plot. A 3-dimensional principal component analysis plot of the data that shows the trends in treatment patterns. Each point represents a patient contact, and the top 3 principal components serve as the axes.

Table 3. Lift Statistics for Treatments Occurring at the Pan American Games (US Personnel Only)

Variable	Cluster			
	I	II	III	IV
Sport				
Archery	0.49[a]	5.05[b]	0.31[a]	0.45
Athletics	0.93	1.84[b]	0.66[a]	0.95
Badminton	0.21[a]	1.03	3.35[b]	0.76
Baseball	0.50[a]	0[a]	1.83[b]	1.82[b]
Basque pelota	1.10	0	0	2.02
Beach volleyball	1.00	0.75	1.31	0.90
Bowling	0	2.75	0	2.69
Boxing	0.44	1.65	1.13	1.61
Canoe kayak	1.10	0[a]	2.51[b]	0.22[a]
Cycling	1.48[b]	0.34[a]	0.23[a]	0.99
Diving	0[a]	0.34	3.05[b]	1.68
Equine	0[a]	0	0[a]	4.04[b]
Field hockey	0.70[a]	0.86	2.62[b]	0.45[a]
Greco-Roman wrestling	0.93	1.09	0.30[a]	1.59[b]
Gymnastics	0.60[a]	0.60	1.12	1.85[b]
Handball	0.91	0.57	1.56	0.97
Judo	0.73	0	0	2.69[b]
Karate	0.52[a]	0	1.34	2.11[b]
Open-water swimming	0.55	0	0	3.03[b]
Pentathlon	1.32	0.83	0	1.21
Raquetball	1.17	0.38[a]	0.60	1.28
Roller skating	1.40	0.75	0	1.10
Rowing	0[a]	0	0	4.04[b]
Sailing	1.15	0.58	0.95	0.97
Squash	0.23[a]	4.78[b]	0.69	0.78
Swimming	1.99[b]	0.05[a]	0.07[a]	0.33[a]
Synchronized swimming	0.73	0.83	0.94	1.61[b]
Table tennis	0.14[a]	4.13[b]	0.70	1.26
Taekwondo	0[a]	0.31	1.46	2.84[b]
Tennis	0[a]	0	4.93[b]	0.50
Triathlon	1.30	1.91[b]	0.72	0.21[a]
Water polo	0.70[a]	1.18	1.36	1.20
Water skiing	0.13[a]	1.46	2.65[b]	1.19
Weightlifting	0.64[a]	1.86	1.82[b]	0.65
Wrestling	0.79	1.27	0.58	1.55[b]
Sex				
Female	1.05	0.80[a]	1.09	0.93
Male	0.96	1.17[b]	0.92	1.05
Position				
Athlete	1.03	1.01	1.01	0.93[a]
Coach	0[a]	0.43	1.48	2.76[b]
Staff	0[a]	0.99	0[a]	3.55[b]
Concern				
Below-the-elbow injury	0[a]	0.79	0.87	3.03[b]
Cervical spine injury	0[a]	3.84[b]	1.73[b]	0.92
Hip/thigh injury	0[a]	2.13[b]	3.11[b]	0.77
Illness	0[a]	0[a]	0[a]	4.04[b]
Knee injury	0.02[a]	0.57	2.01[b]	2.28[b]
Lower extremity injury	0[a]	1.97[b]	1.69[b]	1.86[b]
Lumbar spine injury	0[a]	3.66[b]	2.20[b]	0.67[a]
Recovery	2.08[b]	0.03[a]	0.10[a]	0.14[a]
Shoulder/upper arm injury	0[a]	1.23	2.82[b]	1.38[b]
Thoracic spine injury	0[a]	3.89[b]	2.02[b]	0.69
Provider type				
Allopathic physicians	0.10[a]	0.10[a]	0.31[a]	3.58[b]
Athletic trainers	0.95	0.26[a]	1.46[b]	1.12
Doctors of chiropractic	0.06[a]	4.26[b]	2.01[b]	0.40[a]
Massage therapists	2.12[b]	0.02[a]	0.02[a]	0.13[a]
Osteopathic physicians	0[a]	0.22[a]	0.59	3.50[b]
Physical therapists	0.25[a]	0.26[a]	3.37[b]	1.04

Table 3. Continued

Variable	Cluster			
	I	II	III	IV
Treatment				
Bracing	0[a]	1.24	2.26[b]	1.82[b]
Electrical muscular stimulation	0.05[a]	0.19[a]	3.28[b]	1.50[b]
Graston technique	0[a]	2.64[b]	2.52[b]	0.94
Ice	0.08[a]	0.53[a]	2.11[b]	2.12[b]
Manipulation	0.02[a]	6.68[b]	1.02	0[a]
Massage	2.20[b]	0[a]	0[a]	0[a]
Mobilization	0.02[a]	1.56[b]	3.89[b]	0.45[a]
Moist hot pack	0[a]	0	4.93[b]	0.50
NormaTec[c]	0.62[a]	1.82[b]	0.45[a]	1.69[b]
Other	0[a]	0[a]	3.76[b]	1.35
Proprioceptive neuromuscular facilitation	0.73	0	1.88	1.35
Soft tissue mobilization	0.02[a]	3.66[b]	3.10[b]	0[a]
Stretch	0.12[a]	0.97	3.45[b]	0.87
Taping	0.06[a]	0.97	1.95[b]	2.05[b]
Ultrasound	0.05[a]	2.25[b]	2.05[b]	1.38

[a] Indicates lower level of occurrence than expected from the data ($P < .05$).
[b] Indicates higher level of occurrence than expected from the data ($P < .05$).
[c] NormaTec, NormaTec Recovery Systems, Newton Center, MA.

massages were provided in response to recovery concerns (886 contacts).

Cluster II: Paired Manipulation and Mobilization

Cluster II was characterized by the presence of both STM and CM. The core of this cluster comprised 154 contacts with no additional treatment. Emanating from this core was a trail of 83 contacts that added at least 1 additional treatment for a total cluster size of 237 (Figure 2). The mean cluster II contact involved STM, CM, an 18% likelihood of JM, and less than 10% incidence for each remaining treatment. The cluster was overrepresented by males and members of the archery, athletics, squash, table tennis, and triathlon teams. It was characterized by a high incidence of spinal, hip/thigh, and lower extremity injuries. Contacts were overwhelmingly treated by chiropractors (201 contacts). Patient contacts that involved manipulation paired with STM primarily were conducted by chiropractic physicians (201 contacts).

Cluster III: Soft Tissue Therapy

Cluster III included 347 contacts. Its distinguishing feature was the presence of STM or CM but not both; JM often co-occurred in this cluster. Unlike previously discussed clusters, this cluster did not consist of a single core but rather a group of subcores interpreted as follows (Figure 2): 155 contacts involving STM without JM, 139 contacts involving STM with JM, 39 contacts involving CM without JM, and 14 contacts involving CM with JM.

Most contacts (n = 189) included an additional treatment or treatments, explaining why this cluster was more scattered than the other clusters (Figure 2). The mean cluster III contact had an 85% likelihood of involving STM, 44% likelihood of involving JM, 20% likelihood of involving stretching, 18% likelihood of involving ice,

Table 4. Summary Data for Treatment-Modality Clusters Identified from Medical Contacts of US Personnel at the 2011 Pan American Games

Cluster	Identifier	No. of Contacts	Defining Treatment	Secondary Treatments
I	Massage	888	Contacts involved massage.	Massage was almost always the sole treatment. In 31 contacts, at least 1 other treatment was added.
II	Paired manipulation and mobilization	237	Contacts involved both chiropractic manipulation and soft tissue manipulation but never massage.	In 42 contacts, mobilization was included. In 69, at least 1 additional treatment was included.
III	Soft tissue therapy	347	Contacts involved either soft tissue manipulation (294 contacts) or manipulation (54 contacts), sometimes involved mobilization (153 contacts), but never involved massage.	Treatments occasionally added were stretching (n = 68), ice (n = 64), taping (n = 47), electrical muscular stimulation (n = 25), or at least 1 additional treatment (n = 76).
IV	General medical	485	Contacts were less invasive, never involving soft tissue manipulation, manipulation, or massage.	Included 265 contacts with no treatment and 220 contacts with at least 1 treatment. Most frequent treatments were ice (n = 90), taping (n = 69), mobilization (n = 25), and stretching (n = 24).

15% likelihood of involving CM, 14% likelihood of involving taping, and less than 10% incidence of involving each remaining treatment. This cluster covered a wide range of injuries except for those below the elbow. Contacts predominantly were treated by chiropractors (139 contacts), athletic trainers (108 contacts), and physical therapists (58 contacts). The following sports were overrepresented: badminton, baseball, beach volleyball, canoe kayak, diving, field hockey, tennis, water skiing, and weightlifting.

Cluster IV: General Medical

The remaining 485 contacts formed cluster IV. This was a minimally invasive cluster for which 303 contacts involved no treatment modalities. The mean cluster IV contact had a 19% likelihood of involving ice, 14% likelihood of involving taping, and less than 10% incidence of involving each remaining treatment. Contacts never included STM or CM. This cluster included all 173 illness contacts. Many injuries to the extremities (shoulder/upper arm, below the elbow, knee, lower extremity) belonged in
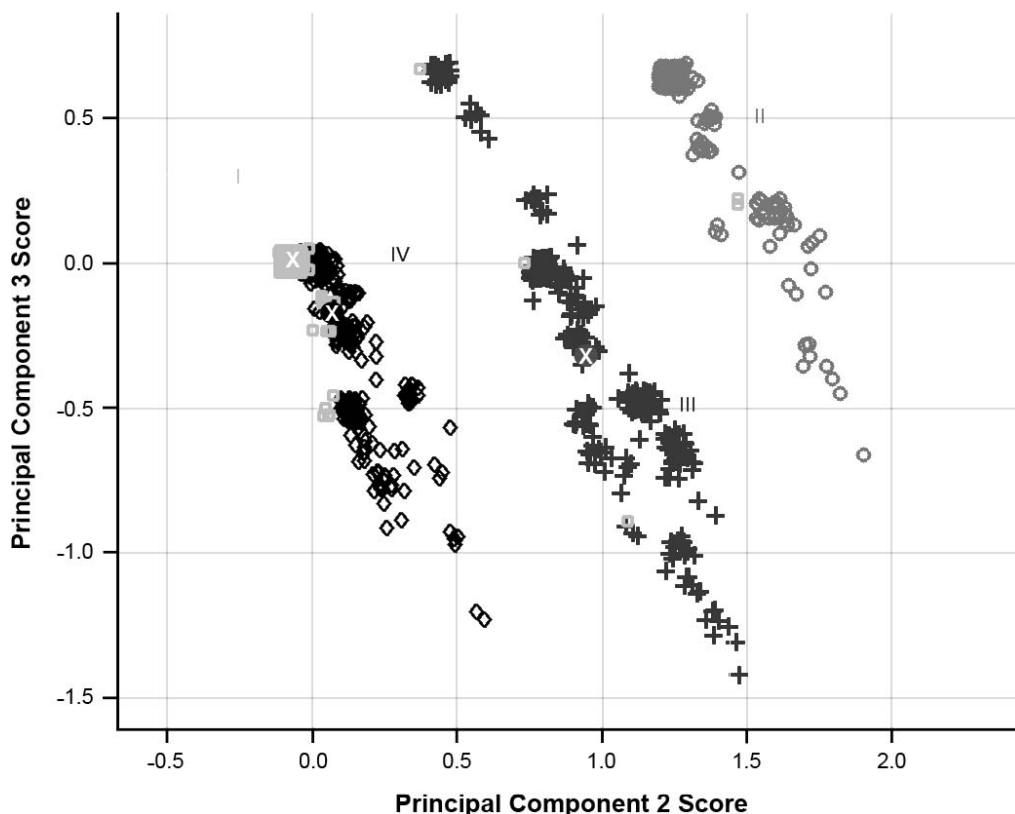

Figure 2. Principal component analysis bi-plot 2-dimensional plot where the second principal component (horizontal axis) is plotted against the third (vertical axis). Each point represents a patient contact. Cluster mean points are represented by a white X.

this cluster. Predominate providers were allopathic (229 contacts) and osteopathic (33 contacts) physicians. Sports that were overrepresented in this cluster were baseball, equine, Greco-Roman wrestling, gymnastics, judo, karate, open-water swimming, rowing, synchronized swimming, and taekwondo.

DISCUSSION

We used the unsupervised technique of PCA to reduce our dataset dimensionality by exploiting correlations among treatments so we could identify trends in 2- and 3-dimensional plots. The major finding was the identification and interpretation of 4 clusters that distinguished distinct treatment modalities or approaches taken by clinicians in treating athletes at the Pan American Games. These modalities are important for assessing treatment methods and determining clinically relevant attributes. Whereas descriptive statistics provide a frequency count of treatments by type, they do not assess underlying treatment patterns that transcend injury or even clinician type. Indeed, the severity of injuries and primary conditions differed across the various sports, yet the underlying treatment pairings that clinicians used held relatively stable at the macro level. In making these associations, we saw very strong links among clusters and provider types. At the most basic level, we observed a strong link between the treatments that team members received and the expertise of the providers. For example, massage therapists tended to provide massages to athletes seeking recovery, and chiropractors tended to offer paired manipulation and mobilization and short tissue therapy for spine-related conditions. These patterns are difficult to identify in a multivariate hyperspace and cannot be visualized using standard descriptive analysis; in our investigation, PCA paired with clustering algorithms easily identified these trends. This observation is particularly interesting, as researchers commonly think that a given treatment protocol is specific to a particular injury or concern. The identified treatment clusters suggested that the clinicians are observing something in practice that drives their treatment to 1 of 4 primary foci. This observation can be exceptionally valuable for national governing bodies that need to ascertain which types of treatments are expected to be used in future competitions and how to best manage equipment and personnel resources. Conversely, the stability in treatment patterns may indicate ingrained preferences of clinicians who do not necessarily consider patient-specific differences or requirements. Regardless of the interpretation, identifying the underlying structure in the data can help clinical directors and national governing bodies strategize better for improvements in patient care at future competitions.

Principal components analysis exploits correlations among variables (in this study, treatments) to discover components that are an efficient linear mapping of the original variables. Trends in a dataset generally allow for a small number of components to capture a high proportion of the relative variance of the data.[15] In our data, many patient contacts were treated in similar ways with paired modalities within a single treatment session, making this dimensionality reduction possible. Therefore, we can find a lower-dimensional representation facilitating the visualization of trends in 2- and 3-dimensional plots.[16]

In this 20-dimensional dataset, 1 trend was relatively easy to identify. The most common treatment (massage) was almost always exclusive of any other treatment. Indeed, 857 contacts involved only a massage, 31 involved a massage and at least 1 other treatment, and the remaining 1069 contacts involved no massage. Other trends were more difficult to spot because they spanned more dimensions than can be visualized. In Figures 2 through 4, components are plotted pairwise. In Figures 3 and 4, it is not obvious that clusters II to IV exist because they appear intermixed, and it is not easy to see in Figure 2 that clusters I and IV are distinct. This highlights the difficulty of searching for patterns in a small number of dimensions and the importance of the tridimensional plot in identifying these 4 clusters (Figure 1).

Traditional analysis would provide the number of massages performed; however, it would not relate the massages to recovery-treatment requests and associate them with the type of clinician performing the treatment. From the current method, we infer that almost half of the treatments administered at the Pan American Games were recovery or preventive and were provided primarily by massage therapists. For future games, these data indicate the need to increase the number of skilled clinicians for the sports teams receiving the most massages to avoid overloading the clinicians assigned to those sports.

These data also illuminated the types of skill sets that are needed for clinicians assigned to specific sport teams. For example, the chiropractic physicians selected for a large international competition need to be experienced with STM techniques for successful delivery of the selected treatments. Advancing this idea, all clinicians traveling with the US team should be provided training before departure on these 4 global treatment modalities to improve communication between providers and facilitate speedy referral of an athlete from 1 provider type to another.

For administrators of the national teams and LOC, these data could also identify preferences of staff physicians or therapists for specific diagnostic tools or imaging devices. Similarly, the analysis can show which sports had a higher use of particular techniques and provide insight into the demands of sports that even an experienced administrator may not be familiar with. For example, Table 3 shows that ultrasound was used 2.25 and 2.05 times more than the average by individuals in the paired manipulation and mobilization (cluster II) and short tissue therapy (cluster III) clusters, respectively. As a whole, these clusters were overrepresented by athletes in the racquet and aquatic sports classifications, with an additional subset of patient contacts from team and multisport athletes and members of the archery and weightlifting teams. Thus, the national governing body should request this device be present or potentially shipped with the other supplies that are to be directed to these specific teams. For the LOC, the data could be paired with scheduling of technicians to ensure adequate availability of the ultrasound for treatment.

Our study had limitations. The analysis was based on a limited number of patient contacts (n = 1957) and features (n = 20). In future studies, researchers should incorporate larger datasets or examine the possibility of amalgamating existing sets of treatment data obtained from large-scale competitions. The primary limitation of this study, however, derives from inconsistency of medical reporting by the various clinicians.
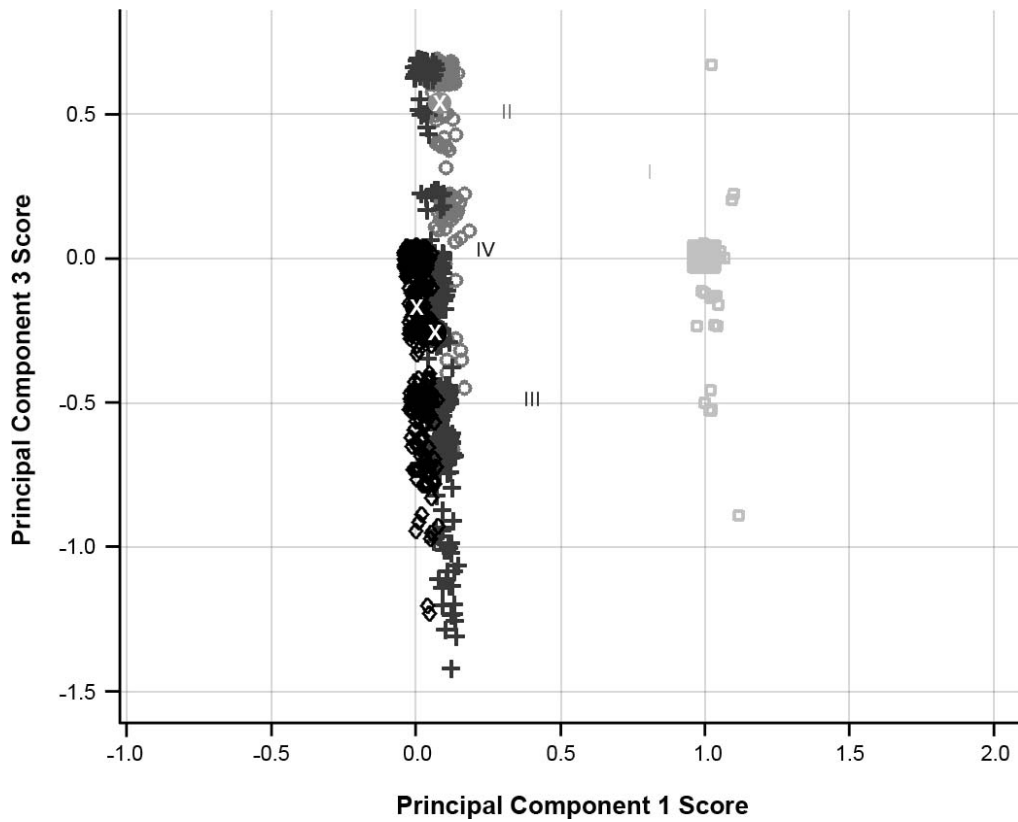
Figure 3. Principal component analysis bi-plot 2-dimensional plot where the first principal component (horizontal axis) is plotted against the third (vertical axis). Each point represents a patient contact. Cluster mean points are represented by a white X.
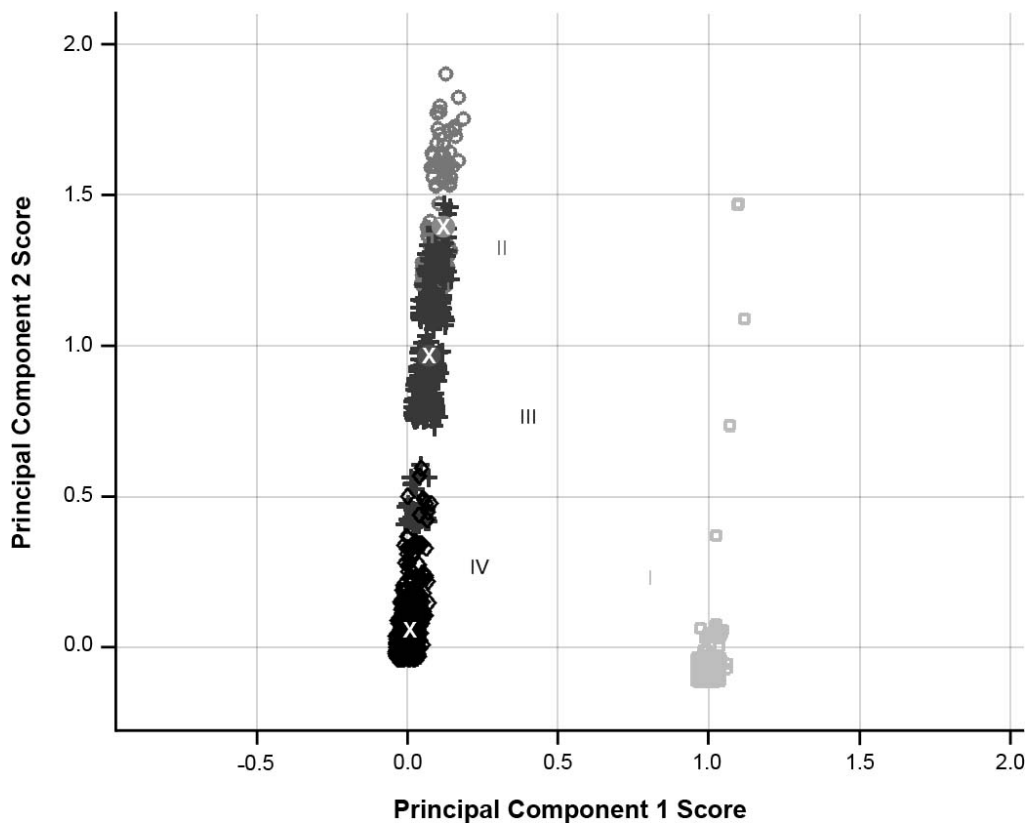

Figure 4. Principal component analysis bi-plot 2-dimensional plot where the first principal component (horizontal axis) is plotted against the second (vertical axis). Each point represents a patient contact. Cluster mean points are represented by a white X.

During the Pan American Games, data were recorded and consolidated manually after the operational period. If this nonrelational system of document collection and transcription is used, an administrator would be limited in the ability to ascertain patterns in the acquired data for improving patient care. The potential breakdown in this type of system is most evident in the failure of some teams to provide patient-contact data. As noted, some teams did not have reported contacts. The effect of these missing data is unclear in our analysis, but the value of this type of modeling is that new information can be incorporated into the existing model, allowing for improved learning as more data become available.

We also noticed during transcription that the outcome data for each patient contact were not explained. This failure was associated with the lack of continuous medical reporting for existing injuries or illnesses that athletes incurred before arriving for competition and to continue treatment that occurred after the athletes returned home. Without an outcome measure, the predictive modeling applications have limited capabilities for these types of datasets. Having a consistent outcome variable to measure patient improvement or deterioration can potentially define ideal treatment patterns for specific injuries with the intent of improving athlete recovery and decreasing the time missed from competition. Given that we only evaluated data that represented a snapshot of the training continuum of the athlete, we recognize that this limitation would require substantial follow-up. Our study, however, represents sufficient proof of concept, showing that unsupervised modeling can illuminate clinically relevant patterns in these types of data.

Other limiting factors were the lack of material consumption during each medical contact and the lack of accounting for the length of each treatment. These data would provide an estimate of sufficient inventory required for treatments and avoid overworking the volunteer sports medicine staff. The review of the treatment numbers does not quantify the associated hours of physical labor provided by each clinician, and most treatments administered were manual therapy techniques. These data would prevent overtasking of clinicians and ensure quality of care from each medical discipline.

CONCLUSIONS

The sports medicine clinic at a large-scale international athletic competition is responsible for providing medical treatment for athletes, coaches, and staff. The multivariate nature of each contact combined with the large number of contacts in general make it difficult to analyze these data efficiently using standard descriptive statistics. Moreover, standard statistical methods do not capture the modality pairings in individual contacts that are characteristic of clinical treatments. We successfully used unsupervised-modeling techniques to identify 4 global treatment modalities that share multiple attributes. Our study provides the USOC and individual national governing bodies with a macroperspective of clinically relevant treatment pairings, as well as a guide for future resource allocation. Whereas this effort was limited to analysis of simple treatment patterns without measures of efficacy, researchers should focus on including an outcome measure to compare prognostic information.

ACKNOWLEDGMENTS

REFERENCES

1. Derman W, Schwellnus M, Jordaan E, et al. Illness and injury in athletes during the competition period at the London 2012 Paralympic Games: development and implementation of a Web-based surveillance system (WEB-IISS) for team medical staff. *Br J Sports Med*. 2013;47(7):420–425.

2. Junge A, Dvorak J, Graf-Baumann T, Peterson L. Football injuries during FIFA tournaments and the Olympic Games, 1998–2001: development and implementation of an injury-reporting system. *Am J Sports Med*. 2004;32(suppl 1):80S–89S.

3. Junge A, Engebretsen L, Alonso JM, et al. Injury surveillance in multi-sport events: the International Olympic Committee approach. *Br J Sports Med*. 2008;42(6):413–421.

4. Engebretsen L, Soligard T, Steffen K, et al. Sports injuries and illnesses during the London Summer Olympic Games 2012. *Br J Sports Med*. 2013;47(7):407–414.

5. Grissom CK, Finnoff JT, Murdock DC, Culberson JT. Nordic venue medical services during the 2002 Winter Olympics. *J Emerg Med*. 2006;30(2):203–210.

6. Rees T. Utah healthcare system watches over Olympians and spectators. *Profiles Healthc Mark*. 2002;18(2):1, 4–5, 3.

7. Dean E. Keeping the games on track. *Nurs Stand*. 2012;26(50):22–24.

8. Bethapudi S, Budgett R, Engebretsen L, O'Connor P. Imaging at London 2012 summer Olympic Games: analysis of demand and distribution of workload. *Br J Sports Med*. 2013;47(13):850–856.

9. Engebretsen L, Steffen K, Alonso JM, et al. Sports injuries and illnesses during the Winter Olympic Games 2010. *Br J Sports Med*. 2010;44(11):772–780.

10. Junge A, Engebretsen L, Mountjoy ML, et al. Sports injuries during the Summer Olympic Games 2008. *Am J Sports Med*. 2009;37(11):2165–2172.

11. Mountjoy M, Junge A, Alonso JM, et al. Sports injuries and illnesses in the 2009 FINA World Championships (aquatics). *Br J Sports Med*. 2010;44(7):522–527.

12. Alonso JM, Junge A, Renstrom P, Engebretsen L, Mountjoy M, Dvorak J. Sports injuries surveillance during the 2007 IAAF World Athletics Championships. *Clin J Sport Med*. 2009;19(1):26–32.

13. Crim JR. Winter sports injuries: the 2002 Winter Olympics experience and a review of the literature. *Magn Reson Imaging Clin N Am*. 2003;11(2):311–321.

14. McDonald J. *Handbook of Biological Statistics*. 2nd ed. Baltimore, MD: Sparky House Publishing; 2009:21–28.

15. Bergeron C, Cheriet F, Ronsky J, Zernicke R, Labelle H. Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression. *Eng Applic Artif Intell*. 2005;18(8):973–983.

16. Demsar U, Harris P, Brunsdon C, Fotheringham AS, McLoone S. Principal component analysis on spatial data: an overview. *Ann Assoc Am Geograph*. 2013;103(1):106–128.

Address correspondence to John P. Vardiman, PhD, LAT, ATC, Department of Human Nutrition, Kansas State University, 210 Justin Hall, Manhattan, KS 66506. Address e-mail to pvardiman@ksu.edu.