© 2020 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2020, 29: 163-176
ISSN 0962-7286
doi: 10.7120/09627286.29.2.163

Theory of medical scoring systems and a practical method to evaluate Asian elephant (Elephas maximus) foot health in European zoos

N Ertl*†‡, P Wendler*‡, E Sós§, M Flügger#, F Schneeweis¶, C Schiffmann‡, J-M Hatt‡ and M Clauss‡

† Hochstraße 14, 95189 Köditz, Germany
‡ Clinic for Zoo Animals, Exotic Pets and Wildlife, Winterthurerstrasse 260, CH-8057, Zurich, Switzerland
§ Budapest Zoo & Botanical Garden, Budapest, Állatkerti Krt 6-12, 1146, Hungary
# Tierpark Hagenbeck, Lokstedter Grenzstraße 2, 22527 Hamburg, Germany
¶ Technische Universität Dresden, Fakultät Mathematik, 01062 Dresden, Germany
* Contacts for correspondence: Nic-Ertl@web.de/paulin.wendler@web.de

Abstract

*Several established models in human and veterinary medicine exist to evaluate an individual health or disease status. Many of these seem unsuitable for further epidemiological research aimed at discovering underlying influential factors. As a case example for score development and choice, the present study analyses different approaches to scoring the foot health of Asian elephants (*Elephas maximus*) living in European facilities. Sum scores with varying degree of detail, and without or with a weighting method, were compared using descriptive statistics, ie kurtosis, skewness, Shannon entropy, total redundancy, their maximum and their actual ranges. With increasing score complexity, a higher level of differentiation was reached. In parallel, the distribution of score frequencies in the population shifted systematically: with the least complex scoring model the pattern indicated a severely unhealthy population with an opposite skew to a hypothetically healthy population, whereas the most complex scoring model indicated a mildly affected population with a skew corresponding to that expected for a healthy population. We propose the latter, in the form of the Particularised Severity Score (ParSev), which accounts for every nail and pad individually and weights the sub-scores by squaring, as the most relevant score for further investigations, either in assessing changes within an elephant population over time, or correlating foot health in epidemiological studies to potentially influencing factors. Our results emphasise the relevance of choosing appropriate scoring models for welfare-associated evaluations, due to implications for the applicability as well as the perceived welfare status of the test population.*

Keywords: *animal welfare, Asian elephant, epidemiology, foot health, scoring system, weighting factor*

Introduction

Foot health of Asian elephants (*Elephas maximus*)

With the elephant being the heaviest terrestrial mammal on the planet, its foot is one of the most important load-bearing structures in the Animal Kingdom. According to a personal communication of Professor DK Lahiri-Choudhury, cited in Csuti *et al* (2001), about 50% of elephants in an Asian working camp are affected by foot problems. Sarma *et al* (2012) came to a similar conclusion with half of their investigated population of Asian elephants in India suffering from foot pathologies, whereas Ramanathan and Mallapur (2008) found that 74.1% of their respective sample population showed pad fissures and 46.9% nail cracks of some description. Under zoo conditions, foot health, especially in Asian elephants (*Elephas maximus*), is a widely discussed and difficult to assess management issue (Csuti *et al* 2001; Fowler 2006). To investigate the *status quo* of Asian elephant foot health in Europe, we determined the prevalence of foot pathologies (Wendler *et al* 2019). Several other studies have investigated links between the prevalence of foot health conditions and husbandry factors (Harris *et al* 2008; Lewis *et al* 2010; Haspeslagh *et al* 2013; Miller *et al* 2016), using different approaches to assess and evaluate foot health status. Due to the differences between those approaches, they depict varying elephant foot health status with prevalence ranging from 67.4 to over 80%. Therefore, meaningful conclusions cannot be readily drawn. For epidemiological evaluations, a quantitative score as an objective measurement of foot health is preferred, yet no commonly accepted method exists as to develop such a score. Here, we present and discuss different approaches to quantify health status in general and their consequences for the perception of a population's health. The Asian elephant population currently living in European zoos presents a suitable example.

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

Evaluating health and disease status

Since the evaluation and prediction of a pathological process is important and, at the same time, somewhat challenging, point-based, risk-scoring models are popular (Austin *et al* 2016). In creating such a model, a series of questions needs to be answered, one of the most important being what method of score calculation to use. One possibility is to follow a 'maximum' concept, by exclusively scoring according to the most severe condition and neglecting all other occurring conditions. For instance, triage scoring systems follow such an approach in cases of having to assess several patients at once in critical situations (Benson *et al* 1996). In such a system, a patient is categorised as 'immediate' and treated without delay, as soon as a pre-defined condition occurs (apnoea or breathing rate > 30 per min or severe bleeding or unconsciousness). A similar 'maximum' concept has been used by the Elephant Welfare Group (N Masters, personal communication 2013). This model assigns the value of its most severe pathology at any location (nail, pad or cuticle) to an elephant, according to a grading system (0–3). In other words, an individual without any lesions except for a single severe one (single sub-score of 3) would be assigned the same total score (3) as an elephant suffering from severe lesions at all possible locations (multiple sub-scores of 3).

Most of the established models in human medicine, however, go for a sum-based evaluation, such as the Glasgow Coma Scale (Jones 1979) or the APGAR score for newborn health (Apgar & James 1962). In these protocols, certain factors are assigned a value, and all values are combined to give a final score, which is used to rank the overall condition. For example, the APGAR score examines respiratory effort, heart rate, muscle tone, skin colour, and reflexes with point values from 0 (bad) to 2 (healthy), leading to a score range from 0 to 10. The newborn is categorised as either 'life at risk' (< 3), 'at risk' (4–6) or 'normal' (> 7). Such a system has, at least theoretically, evident limitations. For instance, there is the theoretical eventuality of a newborn with acute apnoea, but normal values in all other categories and subsequently a score of 8, which would be considered normal, despite life-threatening acute apnoea. With respect to elephant feet, a sum score would sum up the scores given to each individual foot, according to the method applied by Harris *et al* (2008). Similar limitations apply in such a system, as an elephant with three healthy feet (a score of 0) and one foot considered severely affected (a score of 3) would have a lower total score (0 + 0 + 0 + 3 = 3) compared to an elephant with one minor alteration on each foot (1 + 1 + 1 + 1 = 4). In practice, misclassifications due to an atypical distribution of sub-scores may differ in their likelihood between scoring systems, reflecting the inter-dependency of the variables. In the APGAR example, it is extremely unlikely to find an apnoeic newborn with good muscle tone and skin colour. However, in elephants, uneven distributions of pathologies across individual feet appear more frequently (Wendler *et al* 2019).

According to Avila *et al* (2015), a 'formative model' is a concept that consists of several, independently changing, observable factors, as in foot health, which are added up to a final score. Using this approach, a simple sum does not reflect different severities of pathological changes. Therefore, Bollen and Bauldry (2011) or Avila *et al* (2015) emphasise the requirement for a weighting factor in such models. An example that includes a weighting factor is the APACHE model (Acute Physiology and Chronic Health Evaluation), which evaluates certain values of temperature, heart rate, age, and others to predict the likelihood of mortality of a patient (Knaus *et al* 1985, 1991; Zimmerman *et al* 1998). The advantage in developing this model lies in the possibility of verifying the prediction by comparing results with the actual outcome. Thus, it is possible for the revised scores (APACHE II to IV) to adjust weighting factors. Another example is the SAPS model (Simplified Acute Physiology Score) (Le Gall *et al* 1993). In contrast to these models, the introduction of weighting factors appears difficult in a one-time, *status quo*-oriented assessment of elephant foot health without the possibility of evaluating the individual outcome at a later stage.

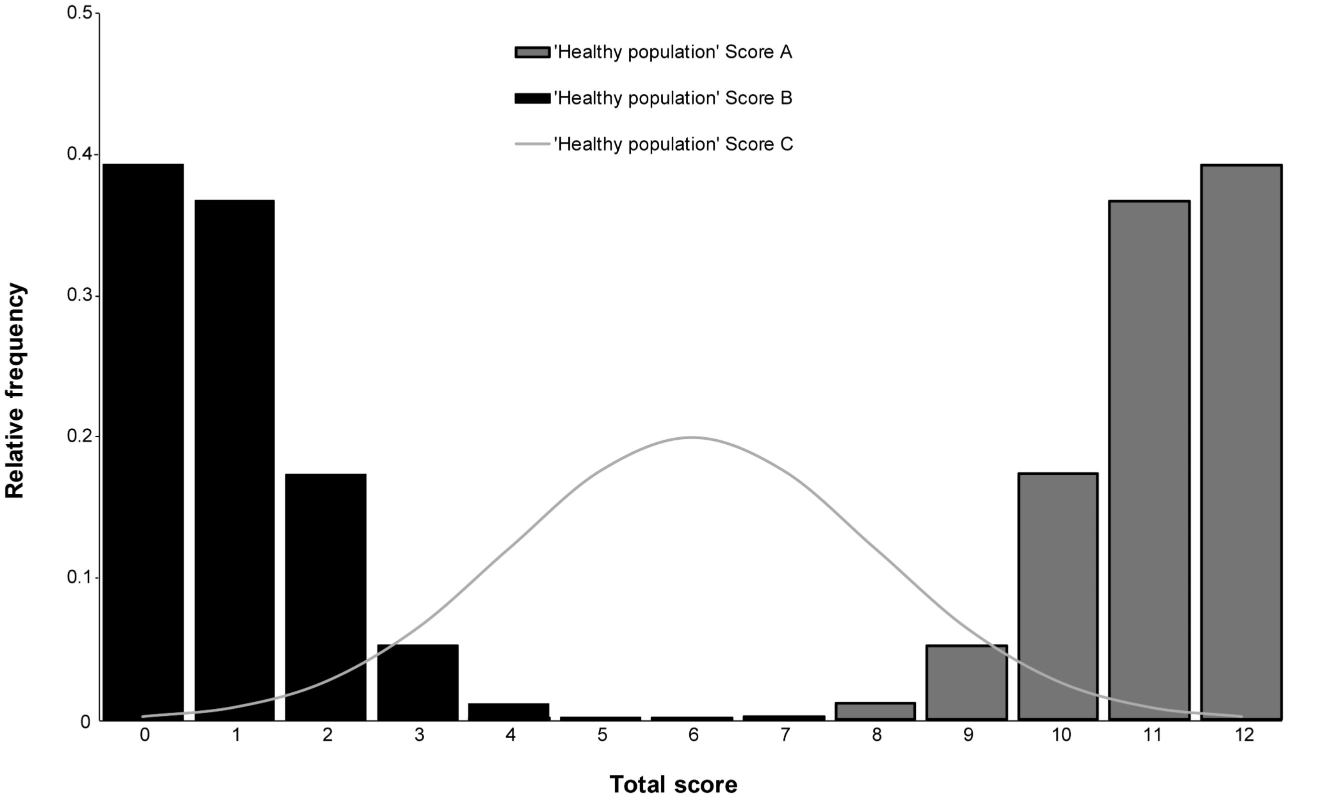
Another important question in developing a score is whether extreme values (low or high) describe a healthy status, or whether the healthy optimum is represented in the middle of the score range. In body condition scores (BCS), the optimum is typically located in the middle of the score range, with both ends being sub-optimal, indicating either cachexia or obesity (Edmonson *et al* 1989). In other systems, certain factors add up to either a healthy status, as in the AGPAR Score (Apgar & James 1962), or a pathological status, as in the score used for foot dermatitis in chickens (Ekstrand *et al* 1994). This results in different expectations for a population's score distributions. In Figure 1, score A would be an example for a model that adds up to a healthy status, as in the APGAR model, with the majority of scored individuals in a relatively healthy sample population showing high values. Score B outlines a model that adds up to a pathological status, as in the foot dermatitis score of chickens. Therefore, the majority of a healthy population has a low score. Score C represents a model where the middle score is favourable, with decreasing numbers of individuals towards low and high scores, displaying a normal distribution. In the case of elephant foot health, resembling a formative model, with several independent components, an approach similar to score B seems appropriate.

In order to represent the actual health status of a population in epidemiological studies, a sufficiently high resolution of a score, which allows distinguishing between mildly and severely affected individuals, is important.

For this purpose, we developed a scoring protocol considering each pathology and all possible locations (each individual nail, each individual pad → 22 locations) similar to the existing foot evaluation of flamingo (*Phoenicoperidae*) feet (Nielsen *et al* 2010; Wyss *et al* 2013). Conditions were classified based on the severity grading of the Elephant Welfare Group's evaluation (N Masters, personal communication 2013) and modified according to Wendler *et al* (2019). Non-pathological care conditions and the pad's surface structure were recorded separately to all pathologies.

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

Figure 1



Theoretical distributions of 'healthy populations' in different scoring models. Score A represents a score where a healthy individual reaches the maximum number of points and deductions are made for health problems. Score B represents a score where a healthy individual has a status of 'zero' and health problems accumulate in the score. Score C represents a score where the optimum is in the middle of the range, with both lower and higher scores indicating non-optimal health conditions.

The intention of this study was to calculate and compare different scoring approaches in assessing an elephant's foot health, in order to determine the best model regarding epidemiological analysis.

Materials and methods

Data collection

Wendler *et al* (2019) investigated the foot health of Asian elephants in 69 institutions registered in the European Endangered Species Programme (EEP). The foot health status of all individuals aged five or older were recorded photographically. This age limit was decided because of the presumed lack of training of animals younger than five years in most institutions. To apply foot-scoring systems, information about all considered structures is necessary and was available for 204 of the examined 243 elephants regarding foot pathologies. For restrictions in training status or enclosure accessibility it was not possible to generate a complete set of photographs for all individual elephants. Evaluation of care status was possible in 191 elephants and of the pads' surface in 222 elephants. The care status was recorded by the use of a care score which sums up the number of non-pathological alterations that can theoretically be removed during a single foot care procedure. Additionally, foot measurements were performed to record the length, width and circumference of each foot, using a soft measuring tape.

Data evaluation

All pathological findings regarding nails and pads were categorised into three grades of severity (1 = mild, 2 = moderate, 3 = severe pathology), whereas healthy structures were scored as 0 (Wendler *et al* 2019). Wendler *et al* describe minor nail cracks and overgrown cuticle as mild, solar horn defects and major nail cracks, as well as fluid pockets in the cuticle and soft tissue areas of the pad as moderate. Purulent discharge of the nail or the pad, altered nail tissue of the cuticle combined with a solar horn defect, and substantial nail lesions are considered the most severe conditions. According to the applied protocol, the rater noted all present pathologies for every location (five nails per front foot, four nails per hind foot, and four pads resulting in 22 locations). The score for each location derived from the worst occurring pathology at this specific location leading to a total of 4^{22} theoretically possible combinations. The resulting data were subsequently interpreted according to a series of scoring protocols.

Calculation of foot health scores

Based on the considerations outlined in the *Introduction*, a 'Maximum Score' was calculated, which attributes the worst scored value of all locations as a total score to an elephant, as noted by N Masters (personal communication

Table 1 Description and calculation of all foot-scoring systems used in the present study.

Score	Description	Formula (range)
Maximum	Total score is the most severe finding in all locations	$\text{score}_{\text{max}}$ [0–3]
Sum	Total score is the sum of the four foot scores	$\text{score}_{\text{LF}} + \text{score}_{\text{RF}} + \text{score}_{\text{LH}} + \text{score}_{\text{RH}}$ [0–12]
Severity	Total score is the sum of the four squared foot scores	$(\text{score}_{\text{LF}})^2 + (\text{score}_{\text{RF}})^2 + (\text{score}_{\text{LH}})^2 + (\text{score}_{\text{RH}})^2$ [0–36]
Particularised Sum	Total score is the sum of all nail and pad scores	$\text{score}_{\text{RFN1}} + \text{score}_{\text{RFN2}}\ldots$ [0–66]
Particularised Severity	Total score is the sum of its separately squared nail and pad scores	$(\text{N1}^2 + \text{N2}^2 + \text{N3}^2 + \text{N4}^2 + \text{N5}^2 + \text{pad}^2)$ for all feet [0–198]
Care	Total score is the sum of all care conditions	[0–62]
Pad	Total score is the sum of the four individual pad scores	$\text{pad}_{\text{LF}} + \text{pad}_{\text{RF}} + \text{pad}_{\text{LH}} + \text{pad}_{\text{RH}}$ [4–16]

Max = maximum; RF = right front foot; LF = left front foot; RH = right hind foot; LH = left hind foot; N = nail.

2013) (range: 0–3). Corresponding to Harris *et al* (2008), a 'Sum Score', based on the maximum sub-scores of the four feet was also evaluated (range: 0–12). Since a limited amount of combinations can reduce the information of a scoring model (Howell *et al* 2007), the number of considered locations was increased for a 'Particularised Sum Score' (ParSum) (range: 0–66) that sums up information from every investigated location (ie not feet, but all nails and pads). In order to avoid the loss of information due to a simple summing up of all sub-scores as mentioned by Avila *et al* (2015), subsequent protocols used squaring as a weighting factor to quantitatively maintain the information that a severity grade of 2 is worse than two severity grades of 1, comparably to the calculation of the Injury Severity Score (ISS) (Baker *et al* 1974). This was done, for every foot's value in the 'Severity Score' (range: 0–36), and again for every location's value in the 'Particularised Severity Score' (ParSev) (range: 0–198) (Table 1). An exemplary calculation for all scores using two fictitious elephants is presented in Table 2.

Additional scores: care and pad score

All conditions that were graded as non-pathological due to the theoretical possibility of being cared for in a single pedicure procedure, were considered as a Care Score by simple addition. It involves three conditions per nail (frayed cuticles, solar fissures, disfigured nail surfaces) and two per pad/foot (frayed pad edges, narrow interdigital spaces between the nails), resulting in a range from 0–62 in an Asian elephant. Those conditions were recorded for subsequent analysis of potential correlations between care status and pathological scores. Since there was a considerable visual difference between the majority of pads, the surface structure of all evaluated pads was considered via a Pad Score which summed up the value of all pads (Wendler *et al* 2019). The single pad's value describes the estimated proportion of so-called 'sulci' or furrows in the surface ($1 < 15\%$; $2 = 15$–29%; $3 = 30$–44%; $4 \geq 45\%$) (range: 4–16) (Table 1). Note that all pads' pathological changes are considered in the foot health scores.

Statistical evaluation

For each of the five foot health scores examined here, the underlying theoretical distribution was calculated, using Matlab R2018a (Moler 1984). This was done under the assumption that all possible 4^{22} individual score combinations occurred with equal frequency and were displayed in all graphs as 'equal distribution'. The actual distributions of the foot health scores were characterised by descriptive statistics (including median and inter-percentile range; skewness, kurtosis and their corresponding 95% confidence intervals; and Kolmogorov Smirnov test for normal distribution).

Skewness describes the extent to which the data distribution resembles a normal distribution with equally diminishing slopes towards the left and right side, or whether the distribution is shifted to one end of the range (Kim 2013). By this definition, a 'sided' score in which the healthy status equals a score of 0 will have a right skew (skewness > 0) if the investigated population is healthy. In Figure 1, the distribution of Score A demonstrates such a right or positive skew. In contrast, score B is negatively or left skewed (skewness < 0).

Kurtosis values describe the position of peaks and outliers compared to a normal distribution. Distributions peaking higher than expected based on a normal distribution have positive kurtosis values (leptokurtic), while negative values indicate an evenly spread ('flat') distribution with less outliers and slopes (platykurtic). For example, if the BCS (in a system ranging from 1–10) of a population showed a very high number of individuals at any particular score (eg an ideal score of 5), with very few individuals having other scores, it would have a positive kurtosis. If, in contrast, scores of 3–7 all occurred at similar frequency in the population, it would have a negative kurtosis. In a 'sided' score, one would expect a high kurtosis if one would assume both a healthy or a particularly unhealthy population.

As a measure of information content and score character redundancy, Shannon entropy and total redundancy were calculated. The Shannon entropy (Shannon *et al* 1949) is used in mathematical communication theory to assess the amount of information per character in a certain data source. It uses the maximal amount and frequency of each available data-point (in our case sub-scores) and results in a number with bits per character as unit. As an example, the Latin alphabet has 26 letters. Due to their asymmetric occurrence, the alphabet shows an entropy of 4.0629 bits per character in contrast to the maximum of 4.7004 (which would result, if all characters appeared equally). For the whole alphabet, this difference can be calculated to a total redundancy of 4.08 characters, ie an alphabet with 22 characters would theoretically suffice for the information typically provided.

A similar approach can help to discover the number of unnecessary characters in scoring models.

To test whether scores show a significant difference to one another, as regards ranking order, Wilcoxon tests were performed, and Spearman rank correlations employed to test the correlation between scores.

Linear foot measurements were regressed against body mass to yield allometric equations in the form of length = aBM^b, with BM = body mass, and an expected geometric exponent of 0.33 (because a length measure should geometrically scale with a volume or mass measure to the power of 0.33) (Clauss & Hummel 2005). These models were calculated as linear regressions after log-transformation (log length = log a + b log BM). We tested whether foot health or care status influenced these allometries by adding the different scores as factors in the regression.

For all statistical calculations R software version 3.4.1. (Ihaka & Gentleman 1993) or SPSS version 23 (IBM 1968) were used. The significance level was set at 0.05.

Results

None of the investigated scores resulted in a normally distributed population. There were significant differences between all scores by Wilcoxon tests ($P < 0.001$), which means that the ranking of animals by their foot health status differed significantly. Despite the notable difference in the ranking of individuals, there were significant correlations between all foot health scores ($P < 0.05$) (Table 3, Figure 2), indicating that the significant difference of the Wilcoxon tests was not caused by an inversion of ranking of individuals between different scoring systems, but by the fact that in the less-differentiated tests, animals had the same score that were further differentiated in the more detailed scoring systems. The Pad Score did not correlate significantly with the Maximum Score, the Particularised Sum Score or the Care Score.

The Maximum Score, the Sum Score, and the Severity Score used their full possible range (suggesting that the worst possible cases actually occurred in the population), whereas the particularised scores did not. Regarding the general distribution of all assigned scores, distinct differences between most of the models were evident. For example, kurtosis values ranged from –0.162 (Particularised Sum Score) to 1.993 (Maximum Score). The health score skewness ranged from a left-skewed distribution of –0.551 (Maximum Score, indicating a population tending towards the 'unhealthy' part of the spectrum) to a clear right-skewed distribution of 1.064 (Particularised Severity Score, indicating a population tending towards the 'healthy' part of the spectrum). Calculated according to the achieved maximum, the Shannon entropy ranged from 1.174 (Maximum) to 5.305 (ParSev). A further computation of total redundancy shows values from 2.086 (70.4%) (ParSum) to 10.446 (5.2%) score characters (ParSev), with the ParSev being the scoring model that used the least amount of available scoring characters (69/198) (Table 4).

Table 2 Exemplary score calculation for two elephants with different foot health status.

Foot	Location	Elephant A	Elephant B
Left front	N1	0	1
	N2	0	1
	N3	0	2
	N4	1	0
	N5	1	0
	Pad	0	0
Right front	N1	1	0
	N2	1	3
	N3	0	2
	N4	0	3
	N5	2	1
	Pad	0	0
Left hind	N2	0	0
	N3	0	0
	N4	2	0
	N5	1	3
	Pad	0	0
Right hind	N2	0	0
	N3	0	0
	N4	0	0
	N5	3	1
	Pad	0	0
Scores			
	Maximum	3	3
	Sum	8	9
	ParSum	12	17
	Severity	18	23
	ParSev	22	39

N = nail; score 0 = no pathology; score 1 = minor pathology; score 2 = moderate pathology; score 3 = severe pathology. Depending on the score model used, the perception of individual health varies. The Maximum and Sum models evaluate elephant A and B as equally affected, whereas the ParSum, the Severity and especially the ParSev models show that elephant B is more severely affected.

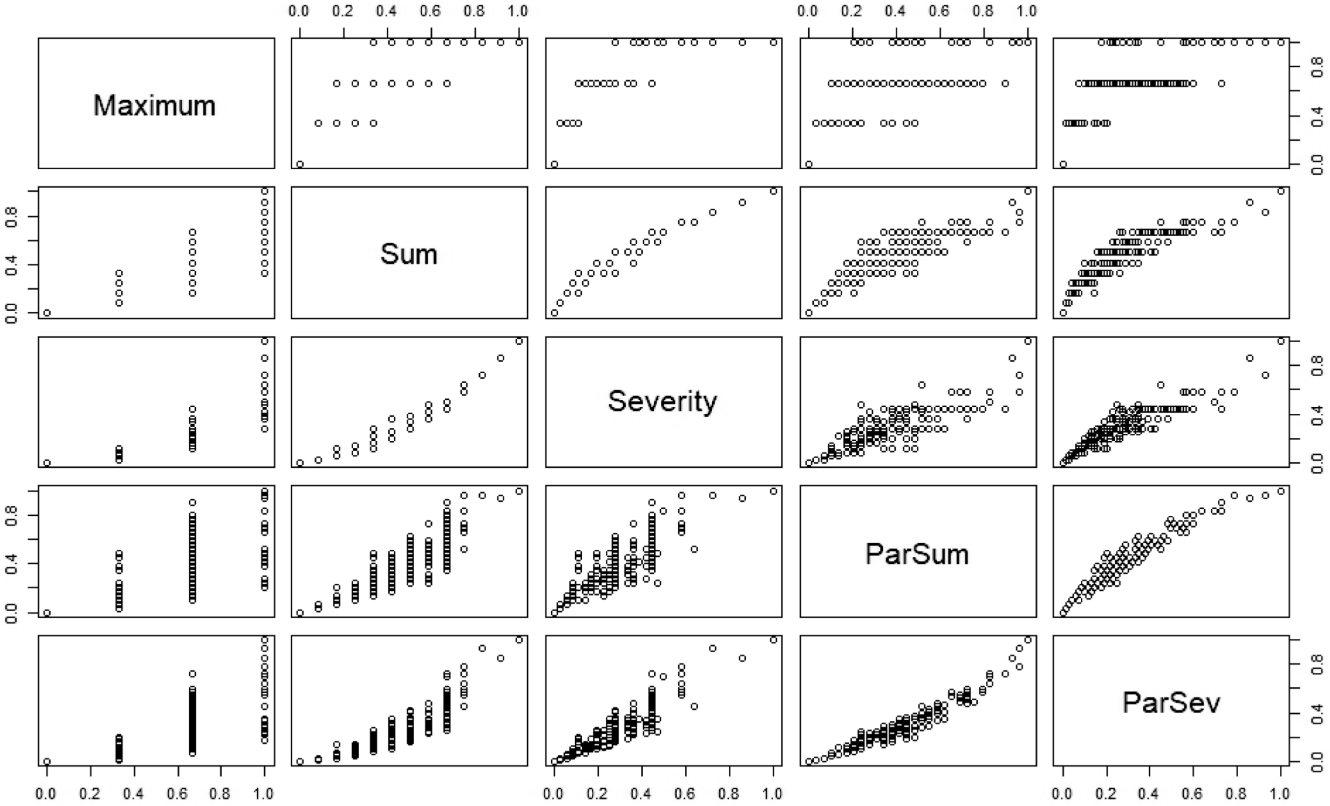
All anatomical measurements met the expectations of a geometric allometric scaling, with an exponent of 0.33 in the 95% confidence interval of the body mass exponent (Table 5). No foot health score had any significant effects on these relationships. The Care and Pad Score, however, were related to length and width allometries, with higher scores associated with higher length or width measures in several cases.

Table 3 Correlation between all elephant foot scores using normalised values.

	Maximum Score	Sum Score	Severity Score	ParSum Score	ParSev Score	Care Score	Pad Score
Maximum Score		$\rho = 0.62$; $P < 0.01$	$\rho = 0.71$; $P < 0.01$	$\rho = 0.49$; $P < 0.01$	$\rho = 0.59$; $P < 0.01$	$\rho = 0.20$; $P < 0.01$	$\rho = 0.13$; $P = 0.057$
Sum Score	$P < 0.01$		$\rho = 0.98$; $P < 0.01$	$\rho = 0.86$; $P < 0.01$	$\rho = 0.92$; $P < 0.01$	$\rho = 0.24$; $P < 0.01$	$\rho = 0.15$; $P = 0.029$
Severity Score	$P < 0.01$	$P < 0.01$		$\rho = 0.81$; $P < 0.01$	$\rho = 0.91$; $P < 0.01$	$\rho = 0.22$; $P < 0.01$	$\rho = 0.16$; $P = 0.022$
ParSum Score	$P < 0.01$	$P < 0.01$	$P < 0.01$		$\rho = 0.96$; $P < 0.01$	$\rho = 0.25$; $P < 0.01$	$\rho = 0.12$; $P = 0.09$
ParSev Score	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$		$\rho = 0.24$; $P < 0.01$	$\rho = 0.16$; $P = 0.019$
Care Score	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$		$\rho = 0.14$; $P = 0.057$
Pad Score	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P < 0.01$	

In combination with Spearman's correlation coefficient ρ (triangle on the right) and results of Wilcoxon tests to compare the ranking of individual animals between two scoring systems (triangle on the left).

Figure 2



Correlation matrix of all scores (normalised to a scale of 0–1) in elephant feet used in the present study. Maximum: Maximum Score that scores an individual according to its worst occurring pathology (0–3); Sum: Sum Score that adds up the four feet score which are in turn scored according to their worst pathology (0–12); Severity: Severity Score that squares the foot values before adding them to weight pathologies (0–36); ParSum: Particularised Sum Score that adds up values from all nails and pads (0–66); ParSev: Particularised Severity Score that squares all nail and pad values before adding them up to weight all pathologies (0–198). Note that individual scores given by a less complex model (eg Maximum and Sum) correspond to a larger number of scores in more differentiated models (eg ParSum and ParSev).

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

Table 4 Descriptive statistics for the different foot-scoring methods.

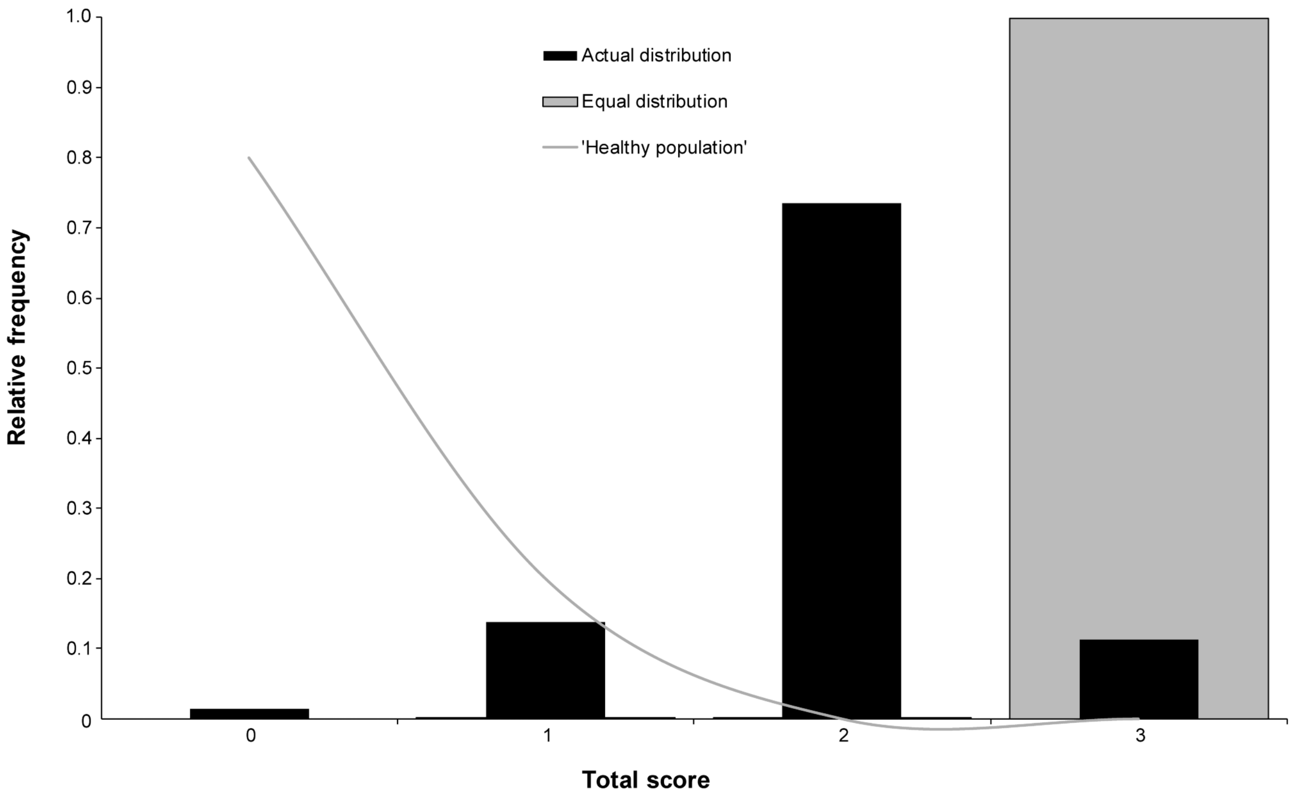
	Maximum Score	Sum Score	Severity Score	ParSum Score	ParSev Score	Care Score	Pad score
Total score range	0–3	0–12	0–36	0–66	0–198	0–62	4–16
N	204	204	204	204	204	191	222
Kolmogorov-Smirnov test	< 0.001	< 0.001	< 0.001	0.003	< 0.001	< 0.001	< 0.001
Median [interpercentile range] (min–max)	2 [0] (0–3)	6 [3] (0–12)	10 [10] (0–36)	11 [8] (0–29)	17 [15] (0–69)	9 [8] (0–30)	10 [6] (4–16)
Kurtosis [CI]	1.993 [1.332–2.660]	–0.136 [–0.800–0.528]	1.589 [0.925–2.253]	–0.162 [–0.826–0.502]	1.615 [0.951–2.279]	0.561 [–0.125–1.247]	–1.078 [–1.715–0.441]
Skewness [CI]	–0.551 [–0.884––0.218]	–0.281 [–0.614–0.052]	0.654 [0.321–0.987]	0.464 [0.131–0.797]	1.064 [0.731–1.397]	0.707 [0.368–1.046]	0.002 [–0.317–0.321]
Shannon entropy [bits/character]	1.174	3.083	3.879	4.532	5.305	4.352	3.657
Total redundancy [character] (% of character range)	2.817 (70.4%)	2.604 (20.0%)	7.322 (19.8%)	2.086 (3.1%)	10.446 (5.2%)	3.824 (6.1%)	0.153 (1.2%)
Summary statement	Severely affected population	Moderately affected population	Moderately affected population	Mildly affected population	Mildly affected population		

Kolmogorov-Smirnov-Test: Tests for normal distribution; Kurtosis: Describes occurrence of outliers in comparison to normal distribution (0 normal distributed; < 0 more evenly distributed than normal; > 0 distribution with higher peaks than normal); Skewness: Describes emphasis of score distribution to the left (right/positive skew; > 0) or to the right (left/negative skew; < 0); Shannon entropy: Describes information content of score character using bits/character as unit. Higher values indicate more information per number; Total redundancy: Describes the number of redundant scores in a model.

Table 5 Scaling relationships of anatomical measurements with body mass according to $y = aBM^b$, with an additional factor c (if significant in log-transformed regression).

y	a [95% CI]	*P*-value	b [95% CI]	*P*-value	R^2	c [95% CI]	*P*-value
Circumference front	10.9 [7.9–15.1]	< 0.001	0.29 [0.25–0.33]	< 0.001	0.72		
Circumference hind	14.8 [11.5–19.2]	< 0.001	0.25 [0.22–0.28]	< 0.001	0.69		
Length front	4.3 [3.1–6.0]	< 0.001	0.27 [0.23–0.31]	< 0.001	0.60		
Length front	4.1 [2.9–5.7]	< 0.001	0.27 [0.23–0.32]	< 0.001	0.63	0.002 [0.001–0.003] (Care score)	0.009
Length hind	5.8 [4.4–7.7]	< 0.001	0.24 [0.21–0.28]	< 0.001	0.64		
Width front	3.5 [2.6–4.8]	< 0.001	0.29 [0.25–0.33]	< 0.001	0.65		
Width front	3.4 [2.5–4.7]	< 0.001	0.29 [0.25–0.33]	< 0.001	0.67	0.001 [0.0–0.003] (Care score)	0.049
Width hind	3.1 [2.1–4.5]	< 0.001	0.27 [0.22–0.31]	< 0.001	0.55		
Width hind	2.8 [1.9–4.0]	< 0.001	0.28 [0.23–0.32]	< 0.001	0.58	0.004 [0.001–0.006] (Pad score)	0.010

Figure 3



Frequency of individual elephant foot health according to the 'Maximum Score' system in different scenarios. Equal distribution assumes that all possible combinations of elephant foot pathologies occur with equal frequency. Actual distribution depicts the results in our sample population. 'Healthy population' describes a hypothetical optimally healthy population. Note the compelling discrepancy between the actual and the hypothetically healthy population, and that an equal occurrence of all possible combinations leads to the impression of a completely unhealthy population.

Discussion

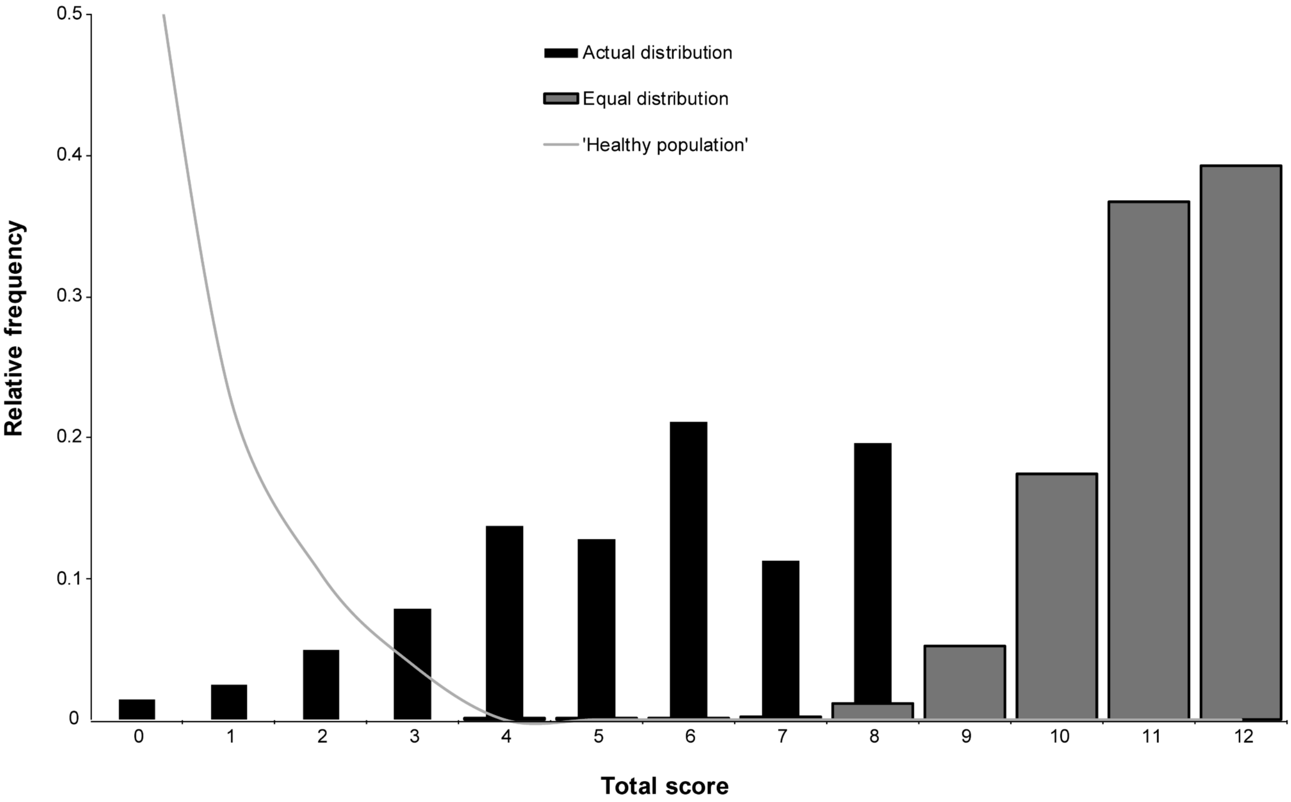
It is worth noting that our scores only describe the current status of foot health in Asian elephants in Europe. There is a need to put this data in context, taking into account potentially influencing factors, such as age or husbandry conditions, but the main aim of the present contribution is a discussion of the effect of designing or choosing a particular scoring system.

Our study demonstrates the challenges of designing an appropriate health score system and ensuring implications for data interpretation. Rules for scoring an individual animal — resembling the typical unit for epidemiological analysis of a population — can lead to drastically different conclusions for the scored population depending on the applied protocol. While our results consistently indicate that the Asian elephant population in Europe shows a certain degree of impaired foot health, the perceived degree varies dramatically between individual scoring systems. The least complex system indicates a severely affected population, with a distribution skewed in the opposite direction to what would be expected for a healthy population, and with a frequency pattern pinpointing virtually equal distribution of each potential combination of pathologies. In contrast, the most complex (ie most differentiated) scoring system displays a mildly affected population, with a distribution skewed towards the direction assumed for a healthy population, and a frequency pattern close to that of a hypothetically healthy population. In addition, the more complex system allows a greater differentiation between individual elephants with a wider spread of sub-scores (0–69), in contrast to the least complex system with scores ranging from 0–3. Moreover, Wilcoxon tests prove a significant difference in ranking order between all scores since scoring systems with fewer sub-scores summarise individuals in the same score that would otherwise vary in ranking order (Figure 2).

The Maximum Score suggests a rather dire health situation. More than two-thirds of all elephants are assigned the second worst total score of two, which results in a negative, left-skewed distribution (–0.551) (Table 4). This distribution creates the impression that most of the sample population is subject to at least moderate pathological changes in their foot health (Figure 3). This is a result of a strong tendency towards higher scores expressed by this protocol, as indicated by its theoretical equal distribution. As a result of the maximum calculation method, the higher scores are by far more likely when assuming an equal distribution than lower scores (score 2: 0.18%, score 3: 99.82%). The actual distribution's kurtosis value of 1.993 hints at a very steep frequency distribution, which is a result of the accumulation

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

Figure 4



Frequency of individual elephant foot health according to the 'Sum Score' system in different scenarios. Equal distribution assumes that all possible combinations of elephant foot pathologies occur with equal frequency. Actual distribution depicts the results in our sample population. 'Healthy population' describes a hypothetical optimally healthy population. Note the stark discrepancy between the actual and the hypothetically healthy population, and that an equal occurrence of all possible combinations leads to the impression of a completely unhealthy population.

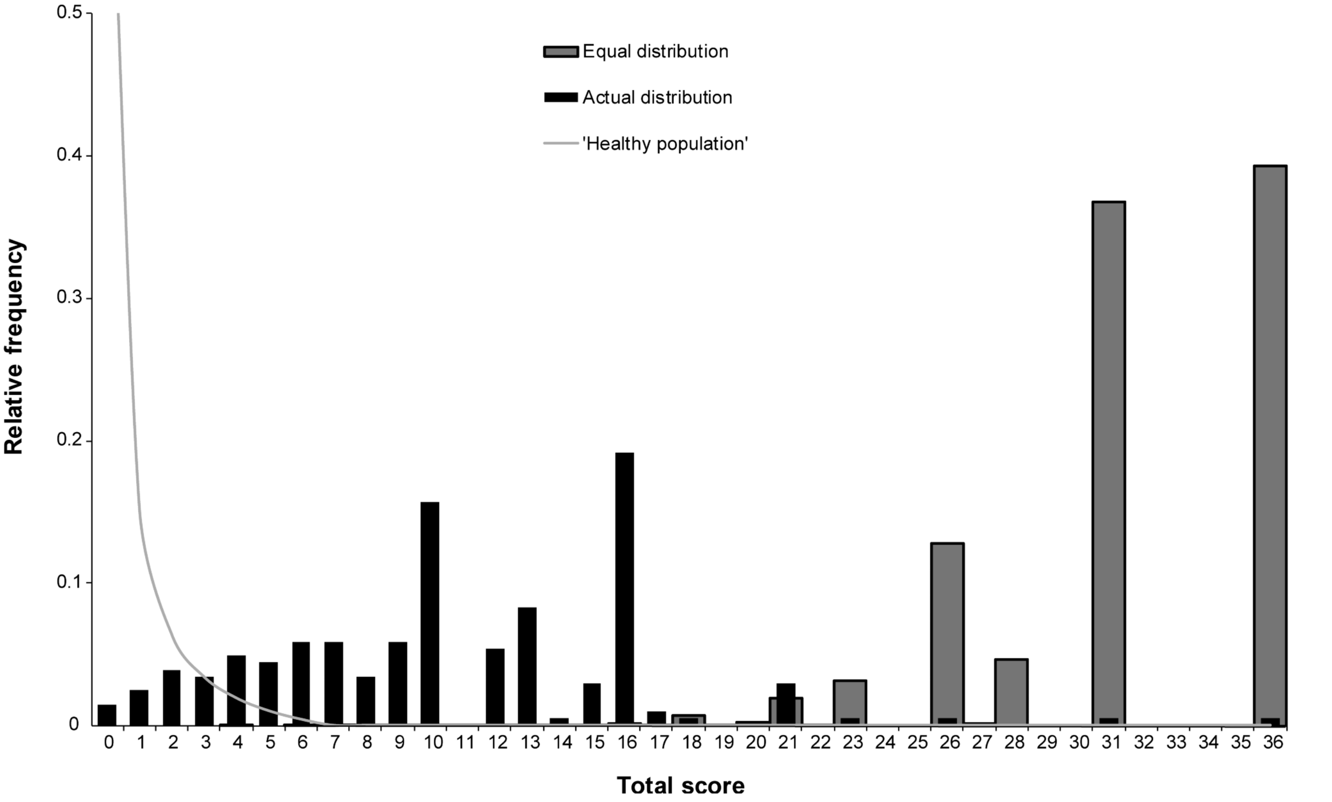
of score 2 individuals. This accumulation also triggers the inter-percentile range of 0, which suggests that most of the scored individuals are assigned with score values extremely close to each other. Shannon entropy indicates that 2.8 characters of the four available are theoretically redundant (ie 70.4% of the score range).

In conclusion, the Maximum Score is completely blurred by its focus on the total score of 2, and for this reason, a rather limited model for our analysis. Restrictions were obvious regarding maximum range, actual range, calculation method and animal-to-animal distinction. We predict this scoring system to have very little value for epidemiological studies on the influence of various factors on foot health.

The slightly more complex Sum Score shows small improvements (Figure 4). Here, the maximum range is reached as well, and a large part of the population is depicted with moderate foot health issues (56.9% with scores over 6). But, as with the Maximum Score, similar calculation limitations exist. In an equal distribution scenario, 98.7% of cases are assigned a value of 9 or higher. However, the added dimension of feet-wise addition pushes the actual distribution towards a more normal one and towards the 'healthy population' at the same time, resulting in a neutral skew of –0.281 with a 95% confidence interval (CI) from –0.614 to 0.052. The same is true for the kurtosis CI of –0.800 to 0.528. An inter-percentile range of 3 showed an increased spread in the single value distribution compared to the maximum model. Compared to the Maximum Score, the Shannon entropy is increased to a value of 3.083 and the redundancy value of 2.604 (20.0% of the score range). Despite these small improvements, this still renders one-fifth of all sub-scores redundant (Table 4).

The Severity Score is characterised by implementing the squaring weight factor for all foot values, which helps to achieve a higher differentiated 'pathological representation'. The equal distribution scenario shows certain restrictions due to the mathematical foundation (Figure 5). Due to the limit of four squarable locations, eight of the 36 score values cannot possibly be computed, and the most frequent combinations (scores over 26: 93.6%) still lead to a left skew in the theoretical distribution. Although the actual distribution shows a shift towards a hypothetical 'healthy population', the maximum range is still reached. The 'squaring peaks' are reflected by a right skew of 0.654, combined with a high kurtosis of 1.589 which describes a high occurrence of outliers compared to the normal distribution. An inter-percentile range of 10 shows a wide spread of sub-scores, being part of the reason why Shannon entropy is increased to

Figure 5



Frequency of individual elephant foot health according to the 'Severity Score' system in different scenarios. Equal distribution assumes that all possible combinations of elephant foot pathologies occur with equal frequency. Actual distribution depicts the results in our sample population. 'Healthy population' describes a hypothetical optimally healthy population. Note the discrepancy between the actual and the hypothetically healthy population, and that an equal occurrence of all possible combinations leads to the impression of a very unhealthy population.

a value of 3.879. The model's calculation limitation becomes evident in a redundancy value of 7.322 (19.8% of the score range), which means that one-fifth of the sub-scores are still redundant, similar to the Sum Score (Table 4).

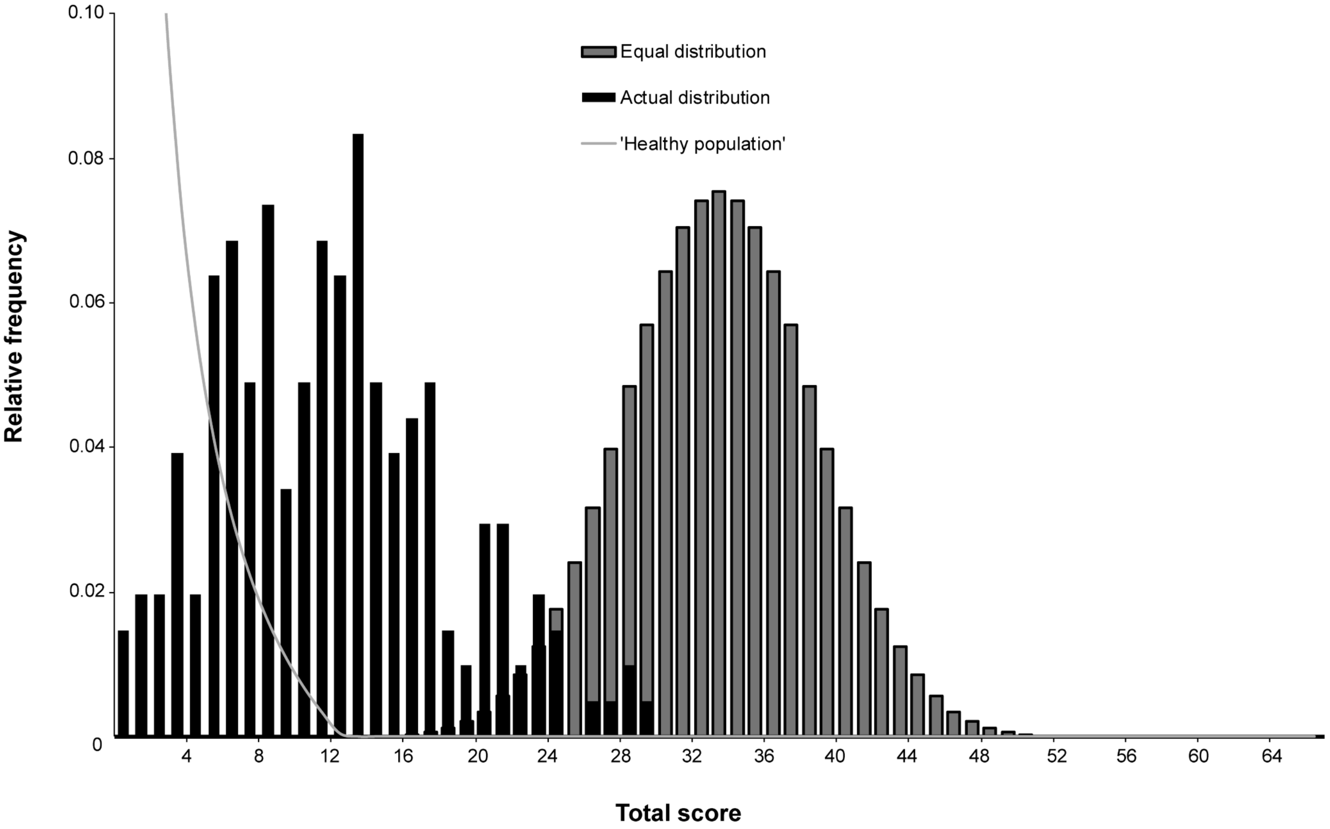
To enhance accuracy, the Particularised Sum Score (Figure 6) considered 22 scoring locations in an Asian elephant. Without the suppressing effect of summarising particular structures by considering only the foot (Sum Score) or even the elephant level (Maximum Score), the equal distribution scenario of this approach shows a well-balanced normal distribution. Since the healthy conditions with scores of 0 are here as likely as pathologies valued with 3, an even curve without any accumulating effect as in prior scoring models is present. The actual distribution indicates a trend towards the theoretically healthy distribution and is therefore right skewed (0.464), but with a low kurtosis of –0.162 (CI: –0826–0.502). The maximum range is not reached (range: 0–29) and the inter-percentile range of 8 shows a fairly even spread of values according to the achieved range. Due to the larger maximum and actual range compared to earlier scores, Shannon entropy is increased to 4.532 and redundancy therefore lowered to 2.086 (3.1% of the score range) (Table 4). Nevertheless, the ParSum Score lacks a weighting factor to stress the severity of moderate and severe lesions.

On the basis of summing every considered location combined with a squaring weighting factor, the ParSev's equal distribution scenario resembles a normal distribution as seen in the ParSum model (Figure 7). The actual distribution shows the highest right skew (1.064) of all analysed scores and again high kurtosis value (1.615) due to the presence of distribution with numerous peaks comparable to the Severity model. Similar to the ParSum model, the maximum range of 198 was not reached (actual range: 0–69) and the occurring sub-scores seem to be relatively evenly spread with an inter-percentile range of 15. The Shannon entropy value of 5.305 shows a further increase in amount of information per character, whereas the ParSev's redundancy is increased (10.446) (Table 4). However, this value corresponds to only 5.2% of the score range.

The analysis of all models showed that the general assessment of a population shifts as scoring models become more detailed and more individual factors (here, nails and pads) are included. Similarly, in the APACHE score development, an addition of more variables from APACHE I with 34 factors to APACHE IV with 142 factors resulted in an additional gain of information (Vincent & Moreno 2010).

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

Figure 6



Frequency of individual elephant foot health according to the 'Particularised Sum Score' system in different scenarios. Equal distribution assumes that all possible combinations of elephant foot pathologies occur with equal frequency. Actual distribution depicts the results in our sample population. 'Healthy population' describes a hypothetical optimally healthy population. Note the actual distribution's shift towards the hypothetically healthy population compared to less complex models, and that an equal occurrence of all possible combinations leads to a normal distribution of score values.

Additional scores

The Care Score showed a low kurtosis of 0.561 (CI: –0.125–1.247) and a right-skewed distribution of 0.707 (Table 4). Furthermore, it did not reach its theoretical maximum range and therefore seemed to describe a relatively well cared-for population. It was felt there was no necessity to assess care conditions employing different severity grades. Consequently, there is no need to implement weighting and it seems appropriate to simply summarise the lack of certain care procedures per elephant. Thus, a Shannon entropy value of 4.352 was found and a relatively low redundancy of 3.824 (6.1% of maximum range). The Care Score was significantly correlated with all foot health scores (Table 3), suggesting that the level of foot care applied to an individual elephant is associated with its foot health status.

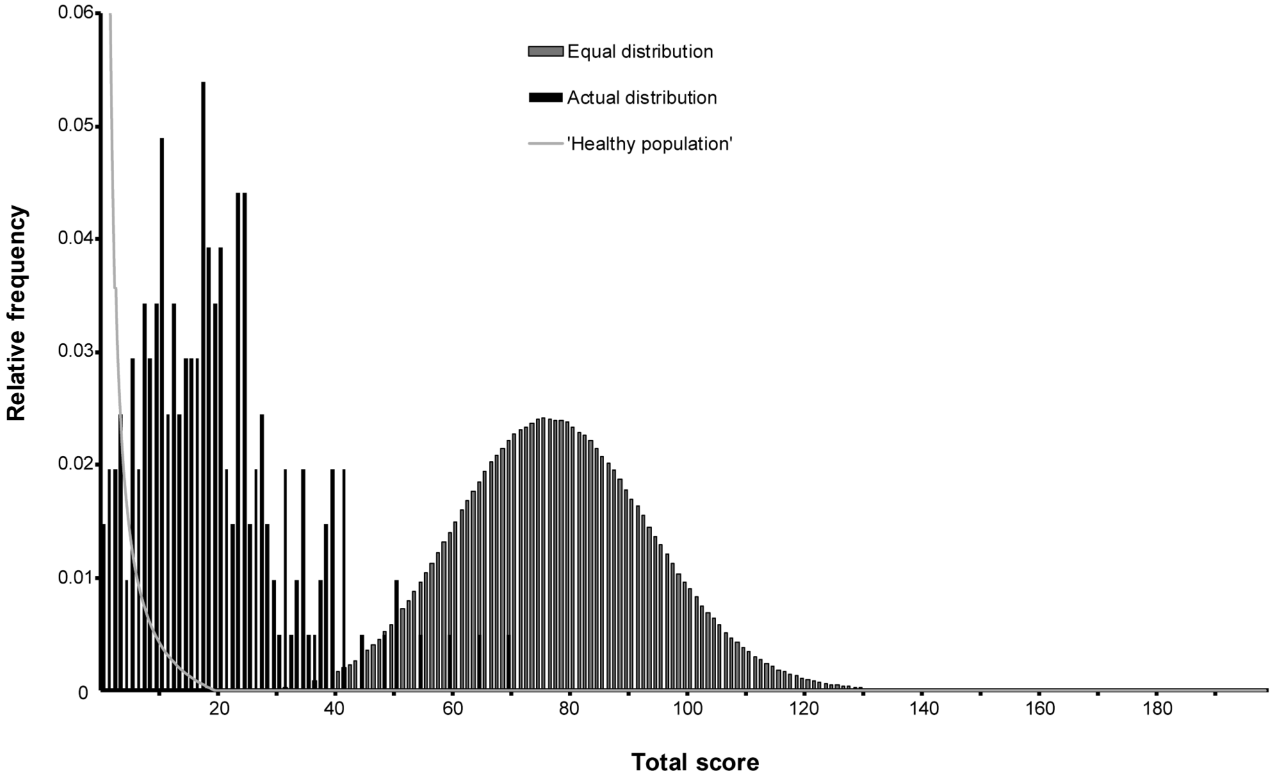
The Pad Score had a strong negative kurtosis of –1.078 and no skewed distribution (0.002). The theoretical maximum was reached and the sub-scores were evenly distributed. The score achieved entropy values of 3.657 and a very low redundancy of 0.153 (1.2% of maximum range) (Table 4). This is the result of the values' even spread without outliers, rendering a very small percentage of characters redundant. A judgment whether any score is more natural or healthy does not seem reasonable, and no emphasis in the distribution is manifest. In particular, there was no significant correlation between the Pad Score and the Care Score or between the Pad Score and two of the five foot-health scores (Table 3). The latter leads to the suggestion that the Pad Score has limited relevance for elephant foot health.

Both non-pathological scores seem to have an influence on the scaling of elephant feet in relation to their body mass (Table 5). This is explained by the fact that less cared for nails and pads tend to be overgrown and thus larger due to the excess skin and nail substance.

Animal welfare implications

More detailed scoring protocols suggest a greater health standard in the investigated population than is indicated by the less-detailed scores, which has implications for the perception of zoo elephant husbandry. Therefore, the choice of a scoring model could also be considered a political one, depending on the agenda of the person or organisation initiating the scoring. In general, applying the model with the highest degree of differentiation seems adequate from a position that aims at understanding a situation in detail. This

Figure 7



Frequency of individual elephant foot health according to the 'Particularised Severity Score' system in different scenarios. Equal distribution assumes that all possible combinations of elephant foot pathologies occur with equal frequency. Actual distribution depicts the results in our sample population. 'Healthy population' describes a hypothetical optimally healthy population. Note the actual distribution's further shift towards the hypothetically healthy population compared to less complex models, and that an equal occurrence of all possible combinations leads to a normal distribution of score values.

holds true until the point of becoming too complex where, even though a larger variety of factors are considered, only limited additional information is gained (Champion *et al* 1980). In the case of Asian elephant foot health in Europe, the ParSev Score is the most robust model, which covers all occurring combination of pathologies. However, it also depicts the zoo elephant population in the most favourable foot health condition compared to other models. This finding is in accordance with the prevalences of individual foot pathologies previously reported for the population under consideration (Wendler *et al* 2019). While 98.5% of all examined elephants showed some form of pathology, only 35.6% of all structures were affected, and only 2.2% of lesions were considered severe. This situation would be poorly reflected by the Maximum Score, which implies a heavily affected population. In conclusion, the ParSev model is a pertinent score to enable an objective analysis of foot health in Asian elephants.

Conclusion

The intention of this study was to calculate and compare different scoring models, as regards their ability to be used as an epidemiological evaluation tool. The most basic Maximum Score model describes a severely affected population whereas the ParSev depicts a very different picture. The implementation of a weighting factor in the most differentiated models allows animals with a few severe lesions to be distinguished from those with many minor pathological changes. We consider this feature practically relevant.

Another important aspect of scoring models is their ability to reflect changes over time. Evidently more differentiated scores are more suited to indicate exacerbation or improvement over time and are recommended when trying to assess effects of modifications to animal husbandry. As Miller *et al* (2016) found it difficult to assess severity and foot problems from veterinary records, our ParSev system provides a numeric value that reflects representative data about an elephant's foot health. This can help to track the foot health development of individual animals and whole populations.

In everyday routine, the model has some disadvantages regarding its overall practicability. Transferring a finding of concern in an elephant into a score is not something required for the management of individual animals, where a detailed description of the foot condition in question and its continuous monitoring and communication in non-abstract terms, is far more important. Scores are more appropriate when seeking epidemiological

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press

status or development surveys of whole populations, for example to assess the average state of welfare, or correlations with other husbandry conditions. While it would be desirable to carry out such surveys on a frequent basis, for example to record the foot health of the European zoo population on a yearly basis and thus monitor development over time, this represents an enormous undertaking that probably cannot be expected to be performed on a routine basis. Most likely, a practical solution is to have certain individuals, such as masters students, perform such surveys at larger time intervals. Since the aim of the foot scores is not to predict a specific outcome, unlike, say, in models for organ function (Multiple Organ Dysfunction Score, Logistic Organ Dysfunction Score or Sepsis-related Organ Failure Assessment Score) (Pettilä *et al* 2002) or patient mortality (APACHE scores), a direct comparison and validation of the accuracy of the scores (to describe a certain outcome) is not feasible. Nevertheless, the model presents a useful tool to quantitatively assess and monitor foot health status of elephants in a cross-sectional as well as longitudinal manner.

Acknowledgements

The authors gratefully acknowledge the co-operation and help of all participating institutions and their staff, the Hagenbeck's foundation for their financial support, and the EAZA Elephant Taxon Advisory Group (TAG) as well as the BIAZA for the study's authorisation. Last, but not least, Paulin Wendler and Nicolas Ertl would like to thank their families for their continuing support and advice in all situations.

References

Apgar VV and James LS 1962 Further observations on the newborn scoring system. *American Journal of Diseases of Children 104*: 419-428. https://doi.org/10.1001/archpedi.1962.02080030421015

Austin PC, Lee DS, D'Agostino RB and Fine JP 2016 Developing points-based risk-scoring systems in the presence of competing risks. *Statistics in Medicine 35*: 4056-4072. https://doi.org/10.1002/sim.6994

Avila ML, Stinson J, Kiss A, Brandao LR, Uleryk E and Feldman BM 2015 A critical review of scoring options for clinical measurement tools. *BMC Research Notes 8*: 612. Ahttps://doi.org/10.1186/s13104-015-1561-6

Baker SP, O'Neill B, Haddon Jr W and Long WB 1974 The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma and Acute Care Surgery 14*: 187-196. https://doi.org/10.1097/00005373-197403000-00001

Benson M, Koenig KL and Schultz CH 1996 Disaster triage: START, then SAVE—a new method of dynamic triage for victims of a catastrophic earthquake. *Prehospital and Disaster Medicine 11*: 117-124. https://doi.org/10.1017/S1049023X0004276X

Bollen KA and Bauldry S 2011 Three Cs in measurement models: causal indicators, composite indicators and covariates. *Psychological Methods 16*: 265-284. https://doi.org/10.1037/a0024448

Champion HR, Sacco WJ, Lepper RL, Atzinger EM, Copes WS and Prall R 1980 An anatomic index of injury severity. *The Journal of Trauma: Injury, Infection and Critical Care 20*: 197-202. https://doi.org/10.1097/00005373-198003000-00001

Clauss M and Hummel J 2005 The digestive performance of mammalian herbivores: why big may not be *that* much better. *Mammal Review 35*: 174-187. https://doi.org/10.1111/j.1365-2907.2005.00062.x

Csuti B, Sargent EL and Bechert US 2001 *The Elephant's Foot: Prevention and Care of Foot Conditions in Captive Asian and African elephants*. Iowa State University Press: Ames, IA, USA. https://doi.org/10.1002/9780470292150

Edmonson A, Lean I, Weaver L, Farver T and Webster G 1989 A body condition scoring chart for Holstein dairy cows. *Journal of Dairy Science 72*: 68-78. https://doi.org/10.3168/jds.S0022-0302(89)79081-0

Ekstrand C, Algers B and Svedberg J 1994 Rearing conditions and foot-pad dermatitis in Swedish broiler chickens. *Preventive Veterinary Medicine 31*: 167-174. https://doi.org/10.1016/S0167-5877(96)01145-2

Fowler M 2006 Foot disorders. In: Mikota SK and Fowler M (eds) *Biology, Medicine, and Surgery of Elephants, Volume One* pp 271-290. Blackwell Publishing Professional: Ames, IA, USA. https://doi.org/10.1002/9780470344484.ch20

Harris M, Sherwin C and Harris S 2008 *The welfare, housing and husbandry of elephants in UK zoos* p 127. University of Bristol: Bristol, UK

Haspeslagh M, Stevens JMG, De Groot E, Dewulf J, Kalmar ID and Moons CPH 2013 A survey of foot problems, stereotypic behaviour and floor type in Asian elephants (*Elephas maximus*) in European zoos. *Animal Welfare 22*: 437-443. https://doi.org/10.7120/09627286.22.4.437

Howell RD, Breivik E and Wilcox JB 2007 Reconsidering formative measurement. *Psychological Methods 12*: 205-218. https://doi.org/10.1037/1082-989X.12.2.205

IBM 1968 *SPSS, 23rd Edition*. IBM: London, UK

Ihaka R and Gentleman R 1993 *R Project, Edition 3.4.1*. The R Foundation for Statistical Computing: Vienna, Austria

Jones C 1979 Glasgow Coma Scale. *The American Journal of Nursing 79*: 1551-1553. https://doi.org/10.2307/3424679

Kim H 2013 Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics 38*: 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Knaus WA, Draper EA, Wagner DP and Zimmerman JE 1985 APACHE II: A severity of disease classification system. *Critical Care Medicine 13*: 818-829. https://doi.org/10.1097/00003246-198510000-00009

Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A and Harrell FE 1991 The APACHE III prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest 100*: 1619-1636. https://doi.org/10.1378/chest.100.6.1619

Le Gall J-R, Lemeshow S and Saulnier F 1993 A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *The Journal of the American Medical Association 270*: 2957-2963. https://doi.org/10.1001/jama.270.24.2957

Lewis KD, Shepherdson DJ, Owens TM and Keele M 2010 A survey of elephant husbandry and foot health in North American zoos. *Zoo Biology 29*: 221-236. https://doi.org/10.1002/zoo.20291

Miller MA, Hogan JN and Meehan CL 2016 Housing and demographic risk factors impacting foot and musculoskeletal health in African elephants (*Loxodonta africana*) and Asian elephants (*Elephas maximus*) in North American zoos. *PLoS One 11*: e0155223. https://doi.org/10.1371/journal.pone.0155223

Moler C 1984 *Matlab, R2018a*. Mathworks, University New Mexico: USA

Nielsen AM, Nielsen SS, King CE and Bertelsen MF 2010 Classification and prevalence of foot lesions in captive flamingos (*Phoenicopteridae*). *Journal of Zoo and Wildlife Medicine 41*: 44-49. https://doi.org/10.1638/2009-0095.1

Pettilä V, Pettila M, Sarna S, Voutilainen P and Takkunen O 2002 Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. *Critical Care Medicine 30*: 1705-1711. https://doi.org/10.1097/00003246-200208000-00005

Ramanathan A and Mallapur A 2008 A visual health assessment of captive Asian elephants (*Elephas maximus*) housed in India. *Journal of Zoo and Wildlife Medicine 39*: 148-154. https://doi.org/10.1638/2007-0008R1.1

Sarma KK, Thomas S, Gogoi D, Sarma M and Sarma DK 2012 Foot diseases in captive elephants. *Intas Polivet 13*: 221-227

Shannon CE, Weaver W and Burks AW 1949 *The Mathematical Theory of Communication*. University of Illinois Press: Illinois, IL, USA

Vincent J-L and Moreno R 2010 Clinical review: scoring systems in the critically ill. *Critical Care 14*: 207. https://doi.org/10.1186/cc8204

Wendler P, Ertl N, Flügger M, Sós E, Schiffmann C, Clauss M and Hatt JM 2019 Foot health of Asian elephants (*Elephas maximus*) in European zoos. *Journal of Zoo and Wildlife Medicine 50(3)*: 513-527. https://doi.org/10.1638/2018-0228

Wyss F, Wenker C, Hoby S, Gardelli B, Studer-Thiersch A, von Houwald F, Schumacher V, Clauss M, Doherr MG, Hafeli W, Furrer S, Bechet A and Robert N 2013 Factors influencing the onset and progression of pododermatitis in captive flamingos (*Phoenicopteridae*). *Schweizer Archiv für Tierheilkunde 155*: 497-503. https://doi.org/10.1024/0036-7281/a000499

Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C and Knaus WA 1998 Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Critical Care Medicine 26*: 1317-1326. https://doi.org/10.1097/00003246-199808000-00012

https://doi.org/10.7120/09627286.29.2.163 Published online by Cambridge University Press