The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2018, 27: 235-249
ISSN 0962-7286
doi: 10.7120/09627286.27.3.235

A review of current indicators of welfare in captive elephants (Loxodonta africana and Elephas maximus)

E Williams*†, CL Chadwick‡, L Yon§ and L Asher#

† School of Animal Rural and Environmental Sciences, Nottingham Trent University, Brackenhurst Campus, Southwell, Nottinghamshire NG25 0QF, UK
‡ Ecosystems and Environment Research Centre, School of Environment and Life Sciences, University of Salford, The Crescent, Greater Manchester M5 4WT, UK
§ School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK
# Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, Framlington Place, Newcastle NE2 4HH, UK
* Contact for correspondence and requests for reprints: ellen.williams@ntu.ac.uk

Abstract

Concerns over elephant welfare in UK zoos have implications for their future in captivity. To monitor improvements made to elephant welfare in UK zoos, non-invasive, valid and reliable indicators of welfare are needed. Using a rapid review strategy and critical appraisal tool, we aimed to appraise evidence from peer-reviewed literature on potential welfare indicators for captive elephants. Scopus, Web of Knowledge and Ovid were searched in January 2014 using terms relevant to captive elephants and welfare assessment. Inclusion and exclusion criteria were applied and remaining articles were critically appraised against a specially designed welfare indicator appraisal tool. Thirty-seven unique indicators of welfare were extracted from 30 peer-reviewed papers which met the inclusion criteria. Behavioural measures of welfare (n = 21) were more common than either physical (n = 11) or physiological (n = 5) measures. Stereotypies were the most frequently used behavioural measure, glucocorticoids were the most frequently used physiological measure and body condition scores were the most frequently used physical measure. There was most support for the following indicators of improved welfare state: reduced stereotypies, reduced glucocorticoids and improved body condition scores. Additional measures which require further validation but had strong associations with the most supported measures, and thus have potential use in welfare assessment, were: increased lying rest and positive social interactions. Further validation of the described measures is needed, but this information forms a crucial part of the knowledge required to efficiently monitor and improve the welfare of elephants in captivity.

Keywords: *animal behaviour, animal welfare, captivity, elephant, welfare assessment, welfare indicators*

Introduction

Zoo elephant welfare across North America and Europe has been publicly criticised in influential reports (Clubb & Mason 2002; Kiiru 2007). These reports have led to a widespread response from animal welfare organisations and the UK Government (Zoos' Forum 2010; Born Free 2015; People for the Ethical Treatment of Animals [PETA] 2015). In 2008, the UK Government called for an independent study with the remit of providing "objective, independent data on the welfare of elephants in the UK" (Harris *et al* 2008). The results from the report by Harris and colleagues (2008) were reviewed by the Zoos' Forum (a government advisory committee) and then used to make recommendations to government. A range of areas of concern were highlighted in the report by Harris and colleagues (2008). In response to these concerns, the Zoos' Forum stated that unless substantial improvements were shown in the health and welfare of captive elephants in the UK and unless there was a compelling reason to breed elephants in the UK, then UK zoos should take steps to stop keeping elephants (Zoos' Forum 2010). In order to document improvements in welfare in any species, including elephants, valid welfare indicators are needed. Here, we use an evidence synthesis approach to identify a suite of welfare indicators for elephants.

For the purposes of this review, animal welfare is considered to be a concept which encompasses both mental and physical health, engagement with the physical or social environment and the opportunity to exhibit control or choice. This is purposefully similar to the definition by Dawkins (2008) who defined welfare as whether or not an animal is healthy and has what it wants.

Maintaining a high level of welfare for animals in any captive environment is of paramount importance. The very nature of the captive environment usually means individuals are exposed to a range of situations which, in all likelihood, they would rarely, if ever, experience in the wild. However, the

assessment of wild animal welfare in captive contexts can be difficult. There are typically few animals of each species in captivity, little standardisation in husbandry and housing (Mason 2010), and limited scope to perform experimental rather than observational studies. Hill and Broom (2009) suggested that the ability of an individual animal to cope with challenges faced in captivity is dependent upon their background and previous experiences. Particularly for elephants, a long-lived species which in captivity have a wide variety of different backgrounds and experiences, measuring individual welfare may be important. Tracking the response of each animal to changes in their environment may allow for measurement of welfare on an individual level.

A number of welfare indicators have previously been identified and used for assessing well-being in captive elephants (for a full review, see Mason & Veasey 2010). The most validated of which were expression of stereotypies (behaviours defined as "repetitive, invariant behaviour patterns with no obvious goal or function"; Mason 1991) and levels of glucocorticoids (GC) (Mason & Veasey 2010). However, it is widely agreed that the use of stereotypies as a sole indicator of welfare must be treated with caution, as if they have become habitual it is likely they are not reliable indicators of current welfare state as they can persist in circumstances that have improved welfare (Mason & Latham 2004). GC must also be interpreted with care as they are an indicator of arousal and thus may be indicative of either positive or negative situations (Ralph & Tilbrook 2016). Furthermore, faecal glucocorticoid metabolites (FGM), which are increasingly used as a non-invasive measure of GC, are confounded by a number of factors, including those directly related to the sample (eg age of sample, collection method) and biological factors (eg sex, age and reproductive status of the animal), which can complicate interpretation (Millspaugh & Washburn 2004). Further suggested measures of welfare in zoo elephants have included skin and foot health, infant mortality rates, signs of affective state and measures of preference or avoidance (Mason & Veasey 2010). When questioned about measures to assess elephant welfare, stakeholders advocated the use of a range of behavioural, physical and physiological indicators of welfare (Chadwick *et al* 2017). The suggested behavioural indicators of good welfare included the presence of natural behaviours (such as social interaction and environmental exploration), lying rest, positive social interactions and behavioural synchrony within groups. Physical indicators of poor welfare which were suggested included being overweight, having poor physical health or being physically unable to lie down. Physiological indicators included GC and temporal gland secretion (Chadwick *et al* 2017). Recent work has investigated the relationship between ovarian cyclicity, prolactin, recumbence, musculoskeletal health, foot health, daily walking distance, body condition score (BCS) and stereotypies and the zoo environment, social life and management (Meehan *et al* 2016). However, despite repeated use of some of these measures in the literature, not all of these measures have been validated for use in welfare assessment.

Hill and Broom (2009) recognised the importance of employing a suite of related measures to attempt to identify the welfare state of an individual animal. A number of papers have assessed one or more behavioural, physical or physiological measurement of elephants in captivity (including zoos, circuses, timber camps) both in their current environment or following changes to their environment or routine (eg loss of a conspecific, change in housing), although the term 'assessment of welfare' was rarely used. Links between the measurements used have occasionally been discussed; however, the reliability and validity of these indicators has never been assessed. Veasey (2006) suggested that documentation of baseline time budgets and comparison with time budgets in new environmental or social conditions, or comparison with wild elephant time budgets may also be a valid means of measuring welfare. Furthermore, being able to reliably predict how a measure of welfare may change following a change of circumstance forms a measure of validity (Meagher 2009).

In order to accurately assess captive elephant welfare through non-invasive measures, it is essential to identify and describe those indicators which provide a reliable and valid assessment of the welfare state of the animal being observed, both at a given time and over a period of time. The indicators should differ between animals in different states of welfare, and results should be repeatable to allow assessment of change over time. In this manuscript, we review and appraise current indictors of welfare which have been applied to individual captive elephants and which have been published in the peer-reviewed literature.

Materials and methods

Search methods: rapid review and critical appraisal

A rapid review (a systematic review which does not include grey literature — books and non-peer-reviewed journal articles in order to provide information in a timely manner) (Harker & Kleijnen 2012) was undertaken in January 2014. Studies were identified and reviewed from searches of 'all years' on the following databases: Scopus, Web of Knowledge (Core Collection, Biosis Citation Index, Biosis Previews, Current Contents Connect, Data Citation Index, Derwent Innovations Index, Medline, Zoological Records [2007–January 2014]) and Ovid (CAB Abstracts, Psycinfo, Zoological Records [1978–2007]). Searches were made of titles, keywords and abstracts during January 2014 using a combination of terms relating to elephants; 'elephant', 'Elephantidae', 'Loxodonta', 'Elephas' and to welfare and husbandry in captivity, eg 'welfare', 'quality of life', 'enrichment', 'husbandry', 'housing', 'behav*', 'stress', 'requirements', 'needs', 'activity', 'movement', 'communication', 'health', 'anticipatory', 'handling', 'drinking', 'eating', 'functional responses', 'living conditions', 'grooming', 'rest', 'antagonis*', 'play', 'repetiti*', 'compulsion', 'self-stimulation', 'posture', 'temperament', 'traits', 'group size', 'psychology', 'learning', 'memory', 'intelligence', 'wellbeing'.

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

Table 1 Definitions used during assessment of reliability and validity of the study methods (based on Meagher 2009; Belshaw *et al* 2015).

Type of reliability or validity	Definition
Intra-rater reliability	The consistency of recording within the same rater repeatedly scoring the same animal within a very short time-frame or ideally the same observation of the animal (as recorded by a video camera)
Inter-rater reliability	The consistency of recording between rates scoring the same animal at the same time or using the same observation of the animal (as recorded by a video camera)
Test re-test reliability	The consistency of answers when scored within the same animal expected to be in the same welfare state after a minimum time interval of two days
Internal reliability	The correlation between items within components of an instrument which are meant to be measuring the same thing. Usually measured with Cronbach's alpha. Allows removal of poor and redundant items during instrument development. Closely related to construct validity
Content/face validity	Whether the items in an instrument appear to be asking what they should be. Logical explanation as to why measure is representative of an aspect of welfare provided in the *Introduction* or *Discussion*
Construct validity	Whether the items in an instrument measure the broad area (construct), which they were designed to measure (eg comfort). Assessed by investigating correlations between similar (convergent) and dissimilar (divergent) welfare measures. These may be other behavioural measures or physiological measures or a combination
Criterion validity (concurrent or predictive)	The results of the instrument are compared to an external, independent criterion measure. The criterion measure is thought to measure the same thing and should ideally be a 'gold standard' test, or an alternative established measure A gold standard measure of welfare could be considered animal choice or strength of motivation either positive or negative; or validated measures of affective state (eg cognitive bias); or (depending on welfare definition) comparison with natural or functional behaviour The criterion measurement is taken from the same animal, and can be at the same time (concurrent to the assessment) or in the future (where the assessment is predictive of the criterion measure) Additional measures of criterion validity would be the ability of the instrument to distinguish between different populations (eg attempt to manipulate welfare, eg provide enriched or impoverished environments and test changes in welfare measures; or compare environments which are believed or previously been shown to have better and worse welfare; or expose to short-term welfare intervention)

Inclusion criteria

Only publications which met all of the following criteria were included in the rapid review and subsequent critical appraisal: (i) Captive elephants (eg those in sanctuaries, zoos, timber camps, circuses, etc) were the main subject of interest or the main focus of the investigation; (ii) the publication contained at least two of the search terms in the abstract; (iii) the publication was available to the authors in full; (iv) the publication was in English; (v) the publication was in a peer-reviewed journal; and (vi) the publication assessed the welfare, behaviour, physiology or physical condition of an elephant at a point in time (or was a proxy for one of these, eg keeper assessment, questionnaires).

Exclusion criteria

Papers were excluded from the rapid review and subsequent critical appraisal if they did not meet all of the above inclusion criteria. Additionally, studies assessing the welfare of captive elephants using methods which could not be applied to an individual (eg retrospective studies assessing population-level reproduction or morbidity rates) were excluded from the review, as these did not fit with the aims of this review. Additionally, measures which involved human interaction (eg keeper-elephant interaction) were not included, due to the complexity of analysis of such a measure. Whilst it is acknowledged that human interaction is an important aspect of welfare, individual differences in keeper-elephant relationships would mean this measure would require more complex analysis, and during this review we were seeking to identify standardised and objective measures which could be universally applied to assess welfare with relative ease.

Application of inclusion and exclusion criteria

A single author (EW) performed the initial database search and applied the inclusion and exclusion criteria to all identified publications. To ensure accuracy and consistency, a random sample of publications from the initial searches (50 papers) were independently assessed using the same inclusion and exclusion criteria by a second author (LA). Information to complete the critical appraisal (Figure 1) was extracted by one author (EW) from all of the publications which met the inclusion criteria. All papers which met all of the inclusion criteria were critically appraised and included in the final review.

Critical appraisal

Each article which met all of the inclusion criteria was critically reviewed to ascertain further details about the study and evaluate the reliability and validity of the work, prior to its inclusion in the review (Table 1). The critical appraisal tool consisted of a series of questions relating to the sample population, the study design, the reliability and validity of the paper, the sampling technique, the method of assessing welfare and the measures of welfare used (Figure 1).

Figure 1

Section 1. INFORMATION ON STUDY POPULATION

1. Species [African (Loxodonta Africana); Asian (Elephas maximus)]
2. Sub-species [L. africana africana; L. africana cyclotis; E. maximus maximus; E. maximus indicus; E. maximus sumatranus]
3. Sample size (males: females)
4. Age range
5. Type of facility or facilities studied [Zoo; Safari park; Circus; Timber Camp; Other]
6. Number of establishments involved in the study
7. Number of enclosures or groups
8. Approximate size(s) of enclosure

Section 2. STUDY DESIGN

9. Study design [*as many as applicable from:* Observational- qualitative ; Observational-quantitative; Retrospective; Prospective; Experimental ; Repeated-measures design; Independent-measures design]
10. Number of repeated measures of same animal
11. Control group used [Yes, No]
12. Study manipulations [Yes, No]
13. Rater blind to study manipulations? [Yes, No]

Section 3. WELFARE INDICATOR METHODS

14. Welfare indicators used *list then complete the remainder of relevant questions for each welfare indicator*
15. Media for data collection [Live observations; Video observations; Proxy assessor (e.g keeper questionnaire); Records]
16. Sampling method [Scan, Focal, Instantaneous, Conspicuous behaviour]
17. Recording method [Continuous; Instantaneous, One-zero]
18. Hours of observations
19. Study time period
20. Time of day of samples (how representative of the time period are the samples) [e.g. Consistent time, spread throughout day, spread throughout night, spread throughout 24 hours, etc]

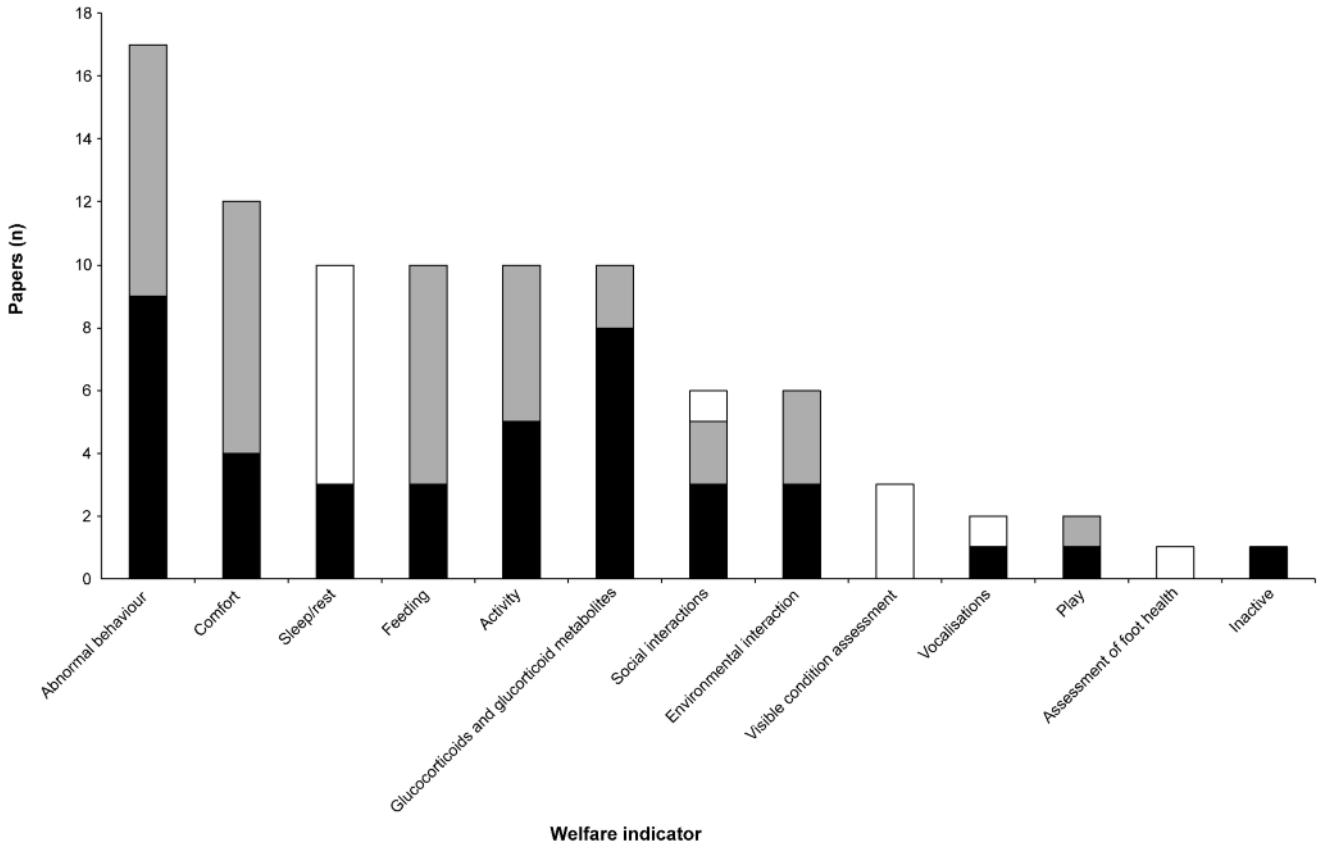
Section 4**. RELIABILITY AND VALIDITY**

21. Types of reliability and validity which have been assessed [*list all that apply and whether they reached criteria for acceptance* Intra-rater reliability; Inter-rater reliability; Test re-test reliability; Internal reliability; Content/face validity; Construct validity; Criterion validity (concurrent or predictive)]
22. Which method was used to assess Criterion Validity of welfare measure? [Presence or absence of motivated items (welfare measured when has and doesn't have items it is motivated to access, approach or avoid); Correlation with behavioural measures of welfare; Correlation with physiological indicators of welfare; Correlation with affective measures of welfare (e.g. cognitive bias); Short term manipulation of welfare state (hours); Long term manipulation of environment (days); Comparison with natural or functional behaviour]
23. What is evidence that criterion validity has been demonstrated? [*For* Comparison with choices and manipulations of welfare state *list Effect size (Mean difference between groups/ standard deviation across groups)*; *For* Correlational designs, *list Correlation coefficients*; *For* Comparison with wild, *list percentage difference in time captive vs wild/ mean time spent in activity in wild*.]
24. Statistics used in the paper [Non-parametric (e.g. Spearman's, correlation, Kruskal Wallis, Wilcoxon); Parametric with no random effects (e.g. Pearson's, t-test, ANOVA, GLM); Modelling or other control for random effects (e.g. mixed models, multi-level)]

Critical Appraisal form for extracting information for critical appraisal of welfare measures used in published studies. Instructions are listed in italics. [Unless otherwise stated mutually exclusive options are listed in square brackets].

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

Figure 2



The total number of papers on captive elephants which conducted studies on each welfare indicator (height of bars), along with the number of papers which have shown a percentage change between treatments presumed to influence welfare (height of grey bars), and the number of these which have demonstrated a significant difference (height of black bars).

Assessment of reliability and validity

Each complete article which met all of the inclusion criteria was assessed for reliability and validity, using questions as detailed in section 4 of the critical appraisal tool (Figure 1), and using pre-defined categories (Table 1). These were independently assessed and recorded for all papers which met the inclusion criteria by two authors (EW, LA) (Table 2; see the supplementary material to papers published in *Animal Welfare* section on the UFAW website: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material); there was no disparity between the authors.

Indicators of welfare

Welfare indicators were extracted from all of the critically reviewed papers and categorised into three broad themes: behavioural, physiological and physical. Within each theme, welfare indicators were grouped as far as possible. A note was made of whether the measures showed: (i) percentage change in concentration or frequency during the period of the study which may not have been subjected to statistical analysis; (ii) statistically significant change in concentration or frequency during the period of the study; (iii) correlation with any other measures of welfare identified during the period of the study; (iv) no change during the period of the study and/or no correlation with any other measures of welfare.

Results

Initial searches yielded 21,000 records, of which 30 publications met all of the inclusion criteria and were critically reviewed. Thirty-seven unique indicators of welfare were extracted from the 30 peer-reviewed papers (for a summary, see Table 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). A complete narrative review of these papers is provided in Appendix 1 (see the supplementary material to papers published in *Animal Welfare* section on the UFAW website: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). The articles were published in 13 different journals, with the majority of articles being found in *Zoo Biology* (eleven papers), *Animal Welfare* (five papers), *Applied Animal Behaviour Science* (three papers) and the *Journal of Applied Animal Welfare Science* (three papers). Nearly half of these studies were conducted by researchers in the USA (14 papers).

Sample size

A summary of the 30 reviewed papers is included in Table 3 (see the supplementary material to papers published in *Animal Welfare* section on the UFAW website: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). The papers reviewed ranged from small, single

elephant, single institution studies to large, multi-institutional studies involving over 100 individuals. The median number of elephants sampled was seven (range 1 to 288). Thirty-one percent of the studies assessed fewer than four elephants, and 9% studied a single elephant. The mean number of institutions included was eight (range 1 to 80). Only 28% of the reviewed papers were multi-institutional studies.

Measures of welfare

Behavioural measures of welfare (21 measures identified) were far more common than either physical (eleven measures identified) or physiological (five measures identified) measures of welfare. Within behavioural measures of welfare, the most frequently used indicators were abnormal (17 papers), comfort (12 papers), feeding (ten papers), locomotion (ten papers), resting (ten papers) and social behaviours (seven papers) (Figure 2). Physical measures of welfare predominantly focused on body condition scoring (three papers). All physiological measurements involved assessment of GC, in particular, faecal glucocorticoid metabolites (FGM) (four papers), salivary cortisol (three papers), and serum cortisol (two papers). An overview of the identified measures of welfare, and whether observed changes were significant, is provided in Table 2 (https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material) and Figure 2.

Behavioural measures of welfare

Behavioural indicators of welfare were broadly separated into nine categories: abnormal, sleep/rest, feeding, environmental interaction, comfort (self-maintenance), activity (walking/locomotion), inactive, social interactions and other (Table 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). Sample sizes ranged from 1 to 140 for behavioural measures of welfare (Table 2) but the majority of studies were based on ten or fewer individuals. Correlation was observed between stereotypical behaviour and five other welfare measures: feeding (negative) (Rees 2009; Koyama *et al* 2012), walking (positive) (Koyama *et al* 2012), resting (negative) (Koyama *et al* 2012), foot health (negative) (Haspeslagh *et al* 2013) and GC (positive) (Wilson *et al* 2004). Sleep/rest and feeding behaviour were both correlated with walking (negative) and stereotypies (negative) (Koyama *et al* 2012). Changes in frequency of social interactions and interactions with the environment were not correlated with any other potential welfare measures, however, associations were identified between increased environmental interaction, reduced stereotypies and increased social interactions in one paper. Frequency of walking or locomotion correlated with rest (negative), feeding (negative) and stereotypic pacing (positive) (Koyama *et al* 2012). Frequency of comfort or self-maintenance behaviours, such as dust-bathing or mud wallowing, were frequently reported in the literature in papers describing general activity budgets of elephants, however, despite being widely reported, comfort behaviours were not correlated with any other measures of welfare. Lesser-used indicators of welfare included inactivity, play behaviour and vocalisations. Correlations between these indicators and more established indicators are yet to be reported.

Physiological indicators of welfare

Measurement of GC and FGM was carried out using various sample types: saliva (three papers), faeces (three papers), serum (three papers) and urine (two papers). Glucocorticoids were noted to correlate with stereotypies (positive) (Wilson *et al* 2004) and specific personality traits (as identified using a keeper assessment of personality): 'fearful' (positive), 'effective' (described as 'gets its own way by controlling other elephants') (negative), 'sociable' (negative) and aggressive (negative) (Grand *et al* 2012; Fanson *et al* 2013). Glucocorticoid assessment was used in studies looking at 1 to 8 elephants (mean 5).

Physical indicators of welfare

The only reported physical indicators of welfare were body condition scores and foot health assessment. All of the study samples for assessment of physical welfare were comparatively large, generally multi-institution studies; body condition was assessed in 82 to 140 elephants (mean 114) in three studies and foot health was assessed in 87 elephants in a single study. Foot health correlated negatively with stereotypies (Haspeslagh *et al* 2013), but otherwise visual assessment of body condition and foot health have not been validated against other behavioural or physiological indicators of welfare.

Reliability and validity of welfare indicators

The overall strength of each measure was assessed based on substantial and biologically meaningful statistical associations with other measures, whether the measure was statistically associated with a previously scientifically validated measure, and the results from the study using the assessed indicator (whether a statistically significant change in the indicator was reported, or a percentage change, and whether this change should be expected based on the conditions experienced by the subject[s] in the study). Indicator strength was also assessed on an individual basis for each study, taking into account the level of validity used by the researchers in the assessment and the number of elephants assessed (Table 3; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). Due to a paucity of information and inconsistency in reporting it was not always possible to garner enough information from the reviewed articles to assess the level of reliability. This information could therefore not be used to assess the strength of the indicator of welfare. Where available, details of test reliability are provided in Table 3. In 15 of the 30 reviewed papers, no assessment of reliability was reported, in five instances measures were taken to increase the reliability of the assessment (eg use of a single observer throughout all observations) but there was no formal statistical assessment, and in ten papers, statistical analysis was undertaken. Level of validity was either explicitly stated or could be ascertained from the information provided, so this information is provided in Table 3; validity reached the construct or criterion level (Table 1) in 26 of the reviewed papers. Levels of reliability were not clear in all of the papers, but in 58% of the reviewed papers there was some form of reliability test detailed (Table 3; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

The reviewed papers assessed welfare over a range of time-periods, using a variety of methods of welfare assessment. Time scales ranged from observations on a small selection of days spread over months or years, to observations in a block of continuous days over a period of days or months. Approaches used to assess welfare included, but were not limited to, monitoring change over time, monitoring change following presumed stressful events, assessing differences between two or more situations, and comparison of the same measures with elephants in the wild. Resting, stereotypies, environmental interaction, feeding, social interactions, self-maintenance behaviours, activity (walking/locomotion) and GC levels all changed significantly when elephants were subject to different environmental or social circumstances. Situations which could be assumed to increase stress and therefore decrease welfare levels, such as transportation, novel flooring and being moved into a smaller enclosure were associated with decreased lying rest, increased standing rest, increased stereotypies, increased GC and decreased environmental interactions. Situations which may be associated with improved welfare, such as being moved into pens/paddocks rather than being chained or shackled, and being provided with time-consuming, naturalistic feeding enrichment were associated with reduced stereotypies, increased feeding, increased positive social interactions, increased self-maintenance and increased activity.

Discussion

The aim of this review was to appraise evidence from current peer-reviewed literature on potential welfare indicators for captive elephants, and to synthesise evidence from the literature on the validity and reliability of these potential welfare indicators. An assessment of the peer-reviewed literature identified a selection of potential welfare indicators for which there was evidence of some level of validity. This included construct and criterion validity for the papers which studied behavioural and physiological indicators, and construct and face validity in the papers studying physical condition. The exact methods of recording each of the welfare indicators varied between studies and therefore any future use of welfare indicators should include assessment of the validity and reliability of the indicator in the context in which it is used (examples of validation processes can be found in Whitham & Wielebnowski 2009 and Wemelsfelder & Mullan 2014). A full narrative review of the welfare indicators is provided in Appendix 1 (https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material); however, it is worth briefly highlighting some of the strengths and limitations of the main welfare indicators identified.

Behavioural indicators

All of the reviewed studies which assessed behavioural indicators of welfare exhibited some degree of criterion validity by a change of state and a further five also exhibited construct validity through statistical association with other welfare indicators. Quantification of the frequency of observed stereotypical behaviour was the most frequently used measure of welfare in the captive elephant literature. Stereotypies are controversial as a welfare indicator because they may not be indicative of current welfare state. Not all stereotypies are sensitive indicators of current welfare state (Mason & Latham 2004); the original factors which caused the stereotypy to develop may not be present in their current environment, and thus a stereotypy may not be a measure of the current welfare of the individual. However, changes in the level of expression of stereotypic behaviour may still be useful as an indicator of welfare when the motivating reasons underlying the performance of the stereotypy are known and when it is coupled with other measures. It has been suggested that an increase in frequency or intensity of stereotypies may be indicative of a welfare issue, and reduction in stereotypies not caused by direct prevention may be indicative of improved welfare (Mason & Latham 2004). The use of stereotypies as an indicator of welfare in the reviewed studies suggests that this is true in these reports; there were meaningful correlations between changes in levels of stereotypical behaviours and other welfare measures. For example, an increase in stereotypies, a decrease in lying rest and an increase in faecal GCM, was observed in a bull elephant post-transport (Laws *et al* 2007). By contrast, a significant decrease in frequency of stereotypical behaviour was observed when elephants were penned rather than chained in a circus (Schmid 1995; Friend & Parker 1999; Gruber *et al* 2000). Formal reliability assessments were reported in seven of the studies and although intra-rater reliability was not assessed, a further two studies used a single observer thereby removing the possibility of inter-rater variation. Used appropriately, ie alongside other suitable measures of welfare and in a situation where there is the opportunity for investigation of change over time, stereotypies appear to be an important and well-supported indicator of welfare. Assessment of stereotypies would be particularly useful to assess an elephant's reaction to changes in housing or husbandry practices; which could then be used to inform management decisions for that elephant.

Although not yet formally validated as an indicator of welfare in elephants, sleep and rest behaviour were linked to other welfare indicators in the reviewed papers, and changed in a predictable manner in a number of different situations. Reliability assessments were conducted in five of the ten papers which assessed sleep behaviour. Reduction in frequency of sleep was correlated with increased stereotypies and associated with events perceived to be stressful to elephants, such as travel (Laws *et al* 2007), death of a conspecific (Koyama *et al* 2012), and introduction of novel flooring (Meller *et al* 2007). Reduced sleep may be indicative of poor welfare in some species, but particularly prolonged periods of time spent asleep may also be indicative of stress (Jones *et al* 2011; McPhee & Carlstead 2012). The quality and pattern of sleep may be important to the welfare of zoo-housed elephants; however, relatively few studies have investigated the resting behaviour of elephants housed in UK zoos (Williams *et al* 2015; Holdgate *et al* 2016b). Elephant keepers and researchers have suggested

that elephants lying down to sleep could be interpreted as indicators of positive welfare, and a lack of sleep or not lying down to sleep could be seen as indicators of negative welfare (Chadwick *et al* 2017). Recent research has shown a relationship between recumbence and substrate, space and social variables in elephants (Holdgate *et al* 2016b) and between some measures of physical health and recumbence (Yon *et al* unpublished) but the complex relationship between rest and recumbence remains unclear. Further research should be undertaken to investigate the factors which affect rest in captive elephants and to investigate the relationship between rest and other welfare indicators, in order to identify if there is an optimal level of lying rest for elephants. However, initial indications suggest that increased lying rest, used in conjunction with other more fully validated measures, could be used as a behavioural measure of welfare in zoo-housed elephants.

A relatively small number of authors researched social interactions in elephants, and social interactions did not correlate with any other welfare measures. However, it is possible that because social interactions were not the main focus of these studies, these less frequently performed behaviours were missed, as these studies focused on compiling activity budgets pre- and post-environmental change. Reliability assessments were undertaken in five of the six reviewed papers. Elephants are a highly social species (Poole & Moss 2008), and reports both in the peer-reviewed literature and by stakeholders suggest that social interactions are an extremely important part of the behavioural repertoire of an elephant. Indeed, in one of the reviewed papers, positive social interactions were greater when elephants were given freedom of choice of social partners and were kept in paddocks rather than being shackled (Schmid 1995), which provided the possibility of more interaction between elephants. It has been suggested by some elephant experts that persistent or extreme aggression within a captive group may be indicative of an underlying welfare problem for either a particular individual or for the entire group (Chadwick *et al* 2017). Other aspects of group behaviour which have been studied in species other than elephants, such as behavioural synchrony (Asher & Collins 2012), or the use of social networks (Asher *et al* 2009), may also be useful welfare indicators. It is felt by stakeholders that social group size is one of the most important factors affecting elephant welfare (Gurusamy *et al* 2014). Used in conjunction with other, validated indicators, expression of positive social interactions should be seen as a positive indicator of welfare.

Walking was widely assessed in the reviewed studies and correlations were observed between rest (negative), feeding (negative) and stereotypic pacing (positive). Five of the reviewed papers investigating walking behaviour formally assessed reliability and a further one used a single observer. Distance elephants travel in the wild has been attributed to availability and distribution of resources (Leighty *et al* 2009); yet, to date, little is known about how far elephants 'should' walk in order to optimise welfare. This study found that elephants housed in larger enclosures and more complex social groups engaged in the greatest amount of walking behaviour (Leighty *et al* 2009), which may be indicative of naturalistic exploratory behaviours. Distance walked has not been found to be related to health or behavioural outcomes, but distance walked has been found to be greater in groups with unpredictable feed schedules and greater number of elephants in the group (Holdgate *et al* 2016a). Individual variability between elephants in walking behaviour within the same environment may be important; a lack of motivation to move, or a physical inability to move owing to poor physical health should be considered as a sign of poor welfare. Walking should be used as an indicator of welfare only alongside other, more traditional indicators, and it should also take into consideration the physical health of the individual elephant and the activities the elephants were engaged with whilst walking, rather than simply distance travelled.

Environmental interactions did not significantly correlate with any other measure; however, increased environmental interaction was associated with positive social interactions and reduced stereotypies. Four of the six reviewed studies assessing environmental interaction and welfare used formal reliability assessments, however, there was disparity between studies in their interpretation and definition of environmental interaction. In order to understand its association with welfare and to increase the validity of this indicator, clear working definitions of activities which constitute environmental interactions must be developed, to enable precision in measuring these behaviours. Environmental interactions could then be used as part of a wider welfare assessment, and if observed in conjunction with other measures, such as reduced stereotypies, environmental interactions may be seen as an indicator of an elephant engaging positively with its environment and therefore experiencing positive welfare.

Not all of the indicators identified in this review necessarily have the immediate potential for welfare assessment, but the presence of species-specific behaviour has been suggested as a potential indicator that the needs of the study animal are being met and that it is experiencing good health and well-being (McPhee & Carlstead 2012). It could therefore be assumed that providing elephants with the opportunity to engage in increased periods of species-typical behaviour is positive for welfare, and that elephants that are engaging in this manner are experiencing good welfare. Species-typical behaviours which require further research before inclusion in welfare assessments for zoo-housed elephants include comfort or maintenance behaviours and feeding. Feeding behaviour correlated negatively with stereotypies and walking in the reviewed studies, however, the factors underlying the relationships between these behaviours are not entirely clear. For example, it is not clear whether it is the lack of opportunity to feed that induces stereotypical behaviour in some elephants, nor is it clear whether the manner of food provision is reducing the distance elephants need to walk. Clubb and Mason (2002) suggested that lack of stimulation to

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

engage in foraging activities is one of the main underlying causes of development of stereotypic behaviour. Researchers have suggested that increased food availability is associated with reduced exhibition of stereotypies (Friend & Parker 1999), and when frequency of foraging is similar to that of wild elephants, relatively little stereotypic behaviour is seen (Koyama *et al* 2012). Indeed, keepers have also suggested that methods of food presentation which enable elephants to engage in more natural feeding behaviours are important for welfare (Chadwick *et al* 2017).

Less recorded but nevertheless important behaviours which have been assessed included play and vocalisations. Further research is needed to investigate these indicators before they can be used reliably in welfare assessment. The small number of studies which have recorded play behaviour may represent the infrequency with which it is recorded in generalised activity budget studies (perhaps due to difficulty defining it), especially in adult elephants, whilst vocalisation data are inherently difficult to capture without specialised recording equipment and require a good knowledge and understanding of the behavioural context for accurate interpretation of the data.

Physiological indicators of welfare

All of the eight reviewed papers on physiological indicators of welfare displayed some degree of either construct or criterion validity; one construct validity only, four criterion validity only and three both construct and criterion validity. Inter-assay reliability assessments were conducted for five of the seven papers. Levels of GC correlated positively with stereotypies and negatively with lying rest. Furthermore, they increased in situations which could be perceived as 'stressful', such as introduction of a new elephant (Dathe *et al* 1992), the opening of the zoo (Menargues *et al* 2008) and transport between facilities (Laws *et al* 2007). Glucocorticoid measurements must be interpreted with caution as an indicator of welfare; GC are produced by the adrenal glands in response to activation of the hypothalamic-pituitary-adrenal (HPA) axis. However, activation of the HPA axis is context-dependent and it may be activated during either beneficial or detrimental circumstances (Palme 2012). Stress responses are an animal's means of coping with their environment (Palme 2012) although it is widely understood that coping mechanisms differ between individuals, and it is not yet clear if there is an 'optimum' coping strategy (Fanson *et al* 2013). Glucocorticoids are also affected by the sex, age, physiological stage, and life history of the animal as well as time of day and environmental factors, such as temperature (for a full review, see Mormede *et al* 2007). Assessment of GC should be used with appropriate consideration of these caveats, measured over a suitable time-period, with a suitable frequency and where possible and appropriate, at a range of time-points throughout the day. They should be investigated in conjunction with a suite of other welfare measures to ensure a complete assessment of welfare.

Physical indicators of welfare

Three papers assessed body condition score of a large number of elephants. These papers only met face validity on our criteria of validity. However, the methods used to assess body condition were designed to increase the accuracy of ratings through thorough assessment, and thus are extremely important when considering the strength of these indicators. An assessment of overall physical condition was achieved through culmination of scores for a number of places on the body, using experienced observers and, in the case of Wemmer and colleagues (2006), designing and trialling the questionnaire using multiple observers and providing pictures and descriptions to increase the accuracy of ratings. Obesity in zoo elephants has been cited as a significant problem, and been linked to poor foot health, arthritis and reduced reproductive output (Clubb *et al* 2008, 2009). Assessment of physical welfare using a body condition scoring protocol has the advantage of being relatively easy to learn and quick to conduct (Wemmer *et al* 2006). Particularly in the captive setting, body condition scoring can be easily incorporated into routine health checks. To investigate the relationship between body condition score and measures of body fat, the method needs to be validated against composition assessments (Wemmer *et al* 2006). However, as a simple means of reliably assessing the overall physical health of an elephant, body condition scores can be an important welfare indicator.

Only one paper included in this review investigated foot health, and that was studied in relation to stereotypies; the study met the threshold level of construct validity. Elephants with higher levels of stereotypies had poorer foot health, but owing to the high percentage of stereotypies observed in the study elephants, the effect could not be identified as causal by the researchers (Haspeslagh 2013). Nevertheless, assessment of foot health is an important physical health indicator as a stand-alone assessment; assessment of physical health, especially foot health, is being increasingly incorporated into preventative care management approaches to keeping elephants in British and Irish zoos (Walter 2010).

Physical indicators of welfare are more likely to change over a longer time scale than behavioural or physiological indicators, making it more challenging to use health measures to assess short-term responses to changing conditions. Furthermore, assessment of some physical welfare indicators, such as foot health, may require closer contact with the animal, so assessment would typically need to be undertaken by animal keeping or care staff, working directly with the elephants, rather than by visiting researchers. However, if undertaken by appropriately trained individuals over time, the methods described in the reviewed papers provide a reliable and valid means of assessing physical welfare of elephants.

Evaluation of the reviewed papers

Welfare assessment models, such as that developed by Sharp and Saunders (2011) utilise systematic, comprehensive and transparent processes to enable evidence-based assessments of animal welfare (Baker *et al* 2016). The finalised assessment must be developed from knowledge of behavioural and physiological responses to changes in circumstance or intervention (Baker *et al* 2016). Such a process was undertaken during this review to identify a suite of potential welfare indicators for use in routine welfare assessment of zoo-housed elephants. As is evidenced in this review, there is a paucity of published literature assessing the welfare of captive elephants; however, there were a number of indicators which have been used repeatedly in the literature which could be used to assess welfare in captive elephants. There were some limitations to the reviewed studies, such as relatively small sample sizes, number of single institution studies (73% of the reviewed papers were single institution studies) and time-period of the studies (80% were conducted in a time-period of less than one year). However, these limitations are, in fact, common to zoo research and not limited to the elephant literature; they arise due to the practical difficulties of conducting long-term, multi-institutional research. It is important not to overlook the importance of numerous single-institution, short-term studies when reviewing the literature, especially when there is relatively little published research. The knowledge gained from these smaller studies could be maximised by using similar or standardised methods and surmising findings across studies. There is also likely to be an intrinsic link between animal-based welfare indicators and environmental conditions (Gurusamy *et al* 2014; Meehan *et al* 2016), so consideration of environmental conditions should be incorporated into future studies with the aim of further validating the identified indicators.

Inclusion of more recently published material

Due to the process required to undertake a systematic review, the findings are only current at the time the search was undertaken. Thus, literature published since the review date may be missed. In this instance, after the review was performed, the results of a large-scale epidemiological study were released, and so it is prudent to include a short review of that work here, even though it was not a formal part of our review. The authors used eight welfare indicators; three behavioural (recumbence, daily walking distance and stereotypy), three physical (BCS, musculoskeletal health and foot health) and two physiological (ovarian acyclity and prolactin levels) (Meehan *et al* 2016). Physical health indicators were associated with situations which may lead to poor welfare. In a study of 255 elephants, a link was established between foot and musculoskeletal health (as measured using presence/absence of abnormalities) and period of time spent on hard surfaces (Miller *et al* 2016). High BCS were prevalent among the studied population of 240 elephants; nearly 75% were considered overweight or obese by the authors (Morfeld *et al* 2016). There was no link between musculoskeletal and foot health and obesity (Miller *et al* 2016) but the authors suggest that management practices which lead to reduced obesity may lead to welfare improvements (Morfeld *et al* 2016). Behavioural measures which were investigated included walking rates and presence of stereotypies. In the 56 elephants studied, distance walked was not related to health or behavioural outcomes. However, walking rates were highest in elephants that had unpredictable feeding schedules and were housed in the largest social groups, and were negatively correlated with overnight space; with elephants having access to larger overnight spaces showing lower walking rates (Holdgate *et al* 2016a). Stereotypical behaviour was the second most prevalent behaviour observed (after feeding) in the study population of 89 elephants (Greco *et al* 2016). The social environment had a significant association with stereotypic behaviour rates: percent time with juveniles and number of elephants housed together contributed to reduced risk of stereotypic behaviour, and being housed separately increased stereotypic risk. However, the authors recognised that there are multiple potential causes which contribute to the expression of stereotypic behaviour (Greco *et al* 2016) and so these effects may not be causal. The final behavioural indicator investigated was recumbence behaviour. Holdgate and colleagues (2016b) studied 72 elephants for, on average, 4 to 5 days each. Species differences were observed between African and Asian elephants; African elephants were recumbent for, on average, 1 h less per day than Asian elephants, and nearly 33% of the studied population were non-recumbent for at least one night. An association was observed between substrate type and recumbence for both species, with resting occurring less frequently on hard flooring. In both species, recumbence was highest in elephants that had the greatest outdoor space overnight. Recumbence was also inversely related to age for both African and Asian elephants — with duration of sleep becoming shorter as elephants aged. Lone-housed elephants slept longer than group-housed individuals, which the authors attributed to a lack of disturbance (Holdgate *et al* 2016b). These findings are in accordance with research by Yon and colleagues (unpublished), which identified less rest in groups with juveniles (due to them disturbing sleep). Yon and colleagues also identified a positive correlation between poor foot health/gait scores and duration of recumbence and found that elephants with poorer physical health were recumbent for longer than average. This highlights the importance of lying rest, but also suggests there may be an optimal level of rest. The relationship between recumbence and welfare remains unclear but it is an important area for further investigation. The findings from this study, using a large number of study animals over a long period of time, contribute to our knowledge of the impact of a number of husbandry factors on elephant welfare. The studies used indicators which are widely used in the literature and support the findings from this rapid review.

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

Evaluation of the critical appraisal tool and methodological limitations

To these authors' knowledge, the critical appraisal tool developed in this paper is the first one developed to review animal welfare measures. This is a significant innovation and one which could be applied to examine welfare measures in other contexts. However, the tool has limitations, some of which relate to adjustments made to the data available. Perhaps the most significant is the use of *P*-values instead of the more informative effect sizes (Nakagawa & Cuthill 2007). When attempting to extract effect sizes we found information provided in most papers did not permit calculation of these values. Due to this lack of information, indicators were classified as either: (i) having a percentage change across different situations; or (ii) having a significant ($P < 0.05$) change across different situations. The questions asked in the critical appraisal tool did not place values on different types of study design or different types of statistics. Other critical appraisal tools consider certain study designs to provide stronger evidence than others, for example, in epidemiological studies, randomised control trials are viewed as the most robust (eg Sibbald & Roland 1998; Kaptchuk 2001; GRADE Working Group 2004). In future developments of this tool, it would be useful to understand the value of different study designs in support of the validity of welfare measures. There are some widely recognised limitations to research conducted on wild animals at captive facilities (see Hosey *et al* 2009). Typically, in critical appraisal, inclusion criteria would be sufficiently stringent that the review would consider only the best quality research with the most appropriate study designs or sample sizes but, in this context, it was not appropriate to be this stringent. If critical appraisal approaches were applied to welfare measures in other species, for which there are fewer limitations on study design and sample sizes, then more selective inclusion criteria should be considered.

The approach used to identify relevant literature also had limitations. Systematic reviews are current only at the point in time at which they are conducted, and thus cannot include work which is published post search date. This review focused only on peer-reviewed literature which was readily available from Scopus, Web of Knowledge and Ovid. Only papers for which the entire article was available were included in the review. It is likely that more evidence exists in 'grey literature', particularly the wealth of information available from within-zoo studies. Such literature is often too inconsistently reported to allow for application of the critical appraisal tool and so was not in the scope of this review, but a narrative review is provided in Asher and colleagues (2015). The papers reviewed ranged widely from single elephant or single institution studies to multi-elephant or multi-institution studies. They also ranged in terms of the level of validity demonstrated for the measures in each study. Many of the reviewed studies did not claim to be 'assessing welfare'; however, if they assessed behavioural change in situations which may be considered to be 'better for welfare' or 'worse for welfare' they were included in the review. The ability to assess change over time is important for an indicator of welfare. However, although we examined the duration of time over which studies were conducted and the methods used to assess welfare, it was beyond the scope of this review to define over what period of time each welfare indicator must be used in order to reliably assess welfare. In future studies, it would be useful to expand the critical appraisal tool to consider the time-period over which welfare indicators were able to detect change. Finally, the biggest constraint when identifying indicators of welfare in captive elephants is that some indicators were more widely used and accepted than others, and these were repeated in the literature. The persistence of the presence of these indicators in the literature does not necessarily indicate that they are the best measures of welfare nor does it mean that they are more useful and should be deemed more important in welfare assessment.

The indicators identified were largely in agreement with welfare indicators suggested by keepers and elephant experts in focus groups (Chadwick *et al* 2017), which may be due to the familiarity many people have with the most common welfare measures. However, there were measures which were mentioned in the focus groups which were not identified in this review of peer-reviewed literature, such as assessment of skin, eyes, gait, and muscle tone. Used in combination, reviews of existing literature and consultation with stakeholders could help identify a range of welfare measures to ensure a complete assessment of welfare for a given species. Indeed, Hill and Broom (2009) suggested that a range of measures must be employed to ensure adequate assessment of welfare in elephants.

Animal welfare implications

Hill and Broom (2009) suggested that the most reliable results come from studies which adopt a multidisciplinary approach to assessing the welfare of animals, ie measuring a wide range of behavioural, physical or physiological indicators. In order to begin to efficiently assess the welfare of captive elephants, a suite of reliable and valid indicators of welfare must be identified. This paper takes the first steps towards identifying and reviewing welfare indicators used previously in the welfare assessment of zoo-housed elephants, synthesising evidence on the reliability and validity of each indicator and identifying from these a selection of behavioural, physical and physiological indicators which could be used in future assessments of captive elephant welfare. This information should be used alongside consultation with zoo staff and other relevant stakeholders, in order to utilise existing knowledge and experience not contained with the scientific literature to identify further possible welfare measures. In this report, a range of different types of welfare indicators have been identified for potential use in assessing the welfare of captive elephants. As previously discussed, the further validation of these welfare indicators would enable the development of a more robust and comprehensive tool for determining captive elephant welfare.

Conclusion

Based on this rapid review and critical appraisal of peer-reviewed literature and assessment of reliability and validity of the reported welfare measures, we suggest that there is support for the following welfare indicators of improved welfare state: reduced stereotypies, reduced GC and improved BCS. Additional measures which are yet to be fully validated but were identified as having strong associations with the listed welfare measures and should therefore be more thoroughly investigated (through inclusion in welfare assessments) are increased lying rest and exhibition of positive social interactions. There is not enough evidence at present to include increased environmental interactions and increased activity (or reduced inactivity) into welfare assessment but they would be worthy of further investigation to establish their future use alongside other, more well-established and validated measures. It is important to note that many of these measures represent a cumulative welfare state, rather than the current welfare state. Thus, a suite of these measures should be employed as part of welfare assessment in elephants. Welfare assessments should incorporate both well-established and validated measures, and some of those measures detailed in this report which have not yet been fully assessed or as frequently used, because of their potential to capture important aspects of welfare. The use of these measures together would enable the assessment of reliability and validity of the less frequently used measures for their use as future welfare measures. Welfare assessments should be repeated within an individual for monitoring purposes, both for routine monitoring over time as part of an ongoing assessment, and following management or husbandry changes to assess a possible response to those changes. The evidence synthesis and critical appraisal approach applied here to evaluate welfare measures could be usefully applied to other contexts and species. The next stage in accurately identifying indicators of welfare in captive elephants is the systematic assessment of the reliability and repeatability of the indicators detailed in this report across a range of conditions in captive elephants over time. This could be achieved through multi-institutional, longitudinal studies of a large number of elephants in a range of different conditions using a standard assessment criterion.

Acknowledgements

This research was conducted as part of a study funded by the Department for Environment, Food and Rural Affairs (Defra). We are grateful to members of the Elephant Welfare Group (EWG) Behaviour Sub-group, Dr Samantha Bremner-Harrison, Dr Oliver Burman and Dr Ros Clubb for advice. We are thankful to Douglas Grindlay for his guidance with the review process. We are appreciative of the invaluable guidance and other support from the policy and evidence team at Defra, and their external experts, Matt Hartley and Dr John Eddison.

References

Asher L and Collins L 2012 Assessing synchrony in groups: Are you measuring what you think you are measuring? *Applied Animal Behaviour Science 138(3-4)*: 162-169. https://doi.org/10.1016/j.applanim.2012.02.004

Asher L, Collins LM, Ortiz-Pelaez A, Drewe JA, Nicol CJ and Pfeiffer DU 2009 Recent advances in the analysis of behavioural organization and interpretation as indicators of animal welfare. *Journal of The Royal Society Interface 6(41)*: 1103-1119. https://doi.org/10.1098/rsif.2009.0221

Asher L, Williams E and Yon L 2015 *Developing behavioural indicators, as part of a wider set of indicators, to assess the welfare of elephants in UK zoos. Defra WC1081*. Defra: Bristol, UK

Baker SE, Sharp TM and Macdonald DW 2016 Assessing animal welfare impacts in the management of European rabbits (*Oryctolagus cuniculus*), European moles (*Talpa europaea*) and carrion crows (*Corvus corone*). *PloS One 11(1)*: e0146298. https://doi.org/10.1371/journal.pone.0146298

Belshaw Z, Asher L, Harvey ND and Dean RS 2015 Quality of life assessment in domestic dogs: An evidence-based rapid review. *The Veterinary Journal 206(2)*: 203-212. https://doi.org/10.1016/j.tvjl.2015.07.016

Born Free 2015 *Elephants in UK and European Zoos*. http://www.bornfree.org.uk/campaigns/zoo-check/captive-wildlife-issues/elephants-in-captivity/elephants-in-uk-and-european-zoos/

Chadwick CL, Williams E, Asher L and Yon L 2017 Stakeholders' opinions on assessing the welfare of elephants (*Loxodonta africana* and *Elephas maximus*) housed in UK zoos. *Animal Welfare 26*: 461-472. https://doi.org/10.7120/09627286.26.4.461

Clubb R and Mason G 2002 *A review of the welfare of zoo elephants in Europe*. RSPCA: Horsham, UK

Clubb R, Rowcliffe M, Lee P, Mar KU, Moss C and Mason GJ 2008 Compromised survivorship in zoo elephants. *Science 322(5908)*: 1649-1649. https://doi.org/10.1126/science.1164298

Clubb R, Rowcliffe M, Lee P, Mar KU, Moss C and Mason GJ 2009 Fecundity and population viability in female zoo elephants: problems and possible solutions. *Animal Welfare 18*: 237-247

Dathe HH, Kuckelkorn B and Minnemann D 1992 Salivary cortisol assessment for stress detection in the Asian elephant (*Elephas maximus*): A pilot study. *Zoo Biology 11(4)*: 285-289. https://doi.org/10.1002/zoo.1430110408

Dawkins MS 2008 What is good welfare and how can we achieve it? In: Dawkins MS and Bonney R (eds) *Future of Animal Farming: Renewing the Ancient Contract*. Blackwell Publishing: Malden, MA, USA

Elzanowski A and Sergiel A 2006 Stereotypic behavior of a female Asiatic elephant (*Elephas maximus*) in a zoo. *Journal of Applied Animal Welfare Science 9(3)*: 223-232. https://doi.org/10.1207/s15327604jaws0903_4

Fanson KV, Lynch M, Vogelnest L, Miller G and Keeley T 2013 Response to long-distance relocation in Asian elephants (*Elephas maximus*): monitoring adrenocortical activity via serum, urine, and feces. *European Journal of Wildlife Research 59(5)*: 655-664. https://doi.org/10.1007/s10344-013-0718-7

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

Friend TH 1999 Behavior of picketed circus elephants. *Applied Animal Behaviour Science 62(1)*: 73-88. https://doi.org/10.1016/S0168-1591(98)00210-X

Friend TH and Parker ML 1999 The effect of penning versus picketing on stereotypic behavior of circus elephants. *Applied Animal Behaviour Science 64(3)*: 213-225. https://doi.org/10.1016/S0168-1591(99)00039-8

Godogama WK, Wemmer C and Ratnasooriya WD 1998 The body condition of Sri Lankan domesticated elephants (*Elephas maximus maximus*). *Ceylon Journal of Science (Biological Sciences) 26*: 1-6

GRADE Working Group 2004 Grading quality of evidence and strength of recommendations. *BMJ 328(7454)*: 1490–0. https://doi.org/10.1136/bmj.328.7454.1490

Grand AP, Kuhar CW, Leighty KA, Bettinger TL and Laudenslager ML 2012 Using personality ratings and cortisol to characterize individual differences in African Elephants (*Loxodonta africana*). *Applied Animal Behaviour Science 142(1-2)*: 69-75. https://doi.org/10.1016/j.applanim.2012.09.002

Greco BJ, Meehan CL, Hogan JN, Leighty KA, Mellen J, Mason GJ and Mench JA 2016 The days and nights of zoo elephants: using epidemiology to better understand stereotypic behaviour of African elephants (*Loxodonta africana*) and Asian elephants (*Elephas maximus*) in North American Zoos. *PLoS One 11(7)*: e0144276. https://doi.org/10.1371/journal.pone.0144276

Gruber TM, Friend TH, Gardner JM, Packard JM, Beaver B and Bushong D 2000 Variation in stereotypic behavior related to restraint in circus elephants. *Zoo Biology 19(3)*: 209-221. https://doi.org/10.1002/1098-2361(2000)19:3<209::AID-ZOO4>3.0.CO;2-7

Gurusamy V, Tribe A and Phillips CJC 2014 Identification of major welfare issues for captive elephant husbandry by stakeholders. *Animal Welfare 23*: 11-24. https://doi.org /10.7120/09627286.23.1.011

Harker JH and Kleijnen J 2012 What is a rapid review? A methodological exploration of rapid reviews in health technology assessments. *International Journal of Evidence-Based Healthcare 10(4)*: 397-410. https://doi.org/10.1111/j.1744-1609.2012.00290.x

Harris M, Sherwin C and Harris S 2008 *The welfare, housing and husbandry of elephants in UK zoos. Defra WC05007*. Defra: Bristol, UK

Haspeslagh M, Stevens J, De Groot E, Dewulf J, Kalmar I and Moons C 2013 A survey of foot problems, stereotypic behaviour and floor type in Asian elephants (*Elephas maximus*) in European zoos. *Animal Welfare 22(4)*: 437-443. https://doi.org/10.7120/09627286.22.4.437

Hill SP and Broom DM 2009 Measuring zoo animal welfare: theory and practice. *Zoo Biology 28(6)*: 531-544. https://doi.org/10.1002/zoo.20276

Hnath PT and Yannessa MN 2002 Effects of facility modifications on elephant activity levels. *Animal Keepers Forum 29(10)*: 421-427

Holdgate MR, Meehan CL, Hogan JN, Miller LJ, Rushen J, de Passillé AM, Soltis J, Andrews J and Shepherdson DJ 2016b Recumbence behaviour in zoo elephants: determination of patters and frequency of recumbent rest and associated environmental and social factors. *PLoS One 11(7)*: e0153301. https://doi.org/10.1371/journal.pone.0153301

Holdgate MR, Meehan CL, Hogan JN, Miller LJ, Soltis J, Andrews J and Shepherdson DJ 2016a Walking behaviour of zoo elephants: associations between GPS measured daily walking distances and environmental factors, social factors and welfare indicators. *PLoS One 11(7)*: e0150331. https://doi.org/10.1371/journal.pone.0150331

Hosey GR, Melfi V and Pankhurst S 2009 *Zoo Animals: Behaviour, Management and Welfare*. Oxford University Press: New York, USA

Jones MA, Mason GJ and Pillay N 2011 Correlates of birth origin effects on the development of stereotypic behaviour in striped mice, *Rhabdomys*. *Animal Behaviour 82(1)*: 149-159. https://doi.org/10.1016/j.anbehav.2011.04.010

Kaptchuk TJ 2001 The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology 54(6)*: 541-549. https://doi.org/10.1016/S0895-4356(00)00347-4

Kiiru W 2007 *The sad state of captive elephants in Canada*. Zoo check Canada, Canada

Koyama N, Ueno Y, Eguchi Y, Uetake K and Tanaka T 2012 Effects of daily management changes on behavioral patterns of a solitary female African elephant (*Loxodonta africana*) in a zoo. *Animal Science Journal 83(7)*: 562-570. https://doi.org/10.1111/j.1740-0929.2011.00992.x

Laws N, Ganswindt A, Heistermann M, Harris M, Harris S and Sherwin C 2007 A case study: fecal corticosteroid and behavior as indicators of welfare during relocation of an Asian elephant. *Journal of Applied Animal Welfare Science 10(4)*: 349-358. https://doi.org/10.1080/10888700701555600

Leighty KA, Soltis J, Wesolek CM, Savage A, Mellen J and Lehnhardt J 2009 GPS determination of walking rates in captive African elephants (*Loxodonta africana*). *Zoo Biology 28(1)*: 16-28. https://doi.org/10.1002/zoo.20199

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(2)*: 221-236

Mason GJ 1991 Stereotypies: a critical review. *Animal Behaviour 41(6)*: 1015-1037. https://doi.org/10.1016/S0003-3472(05)80640-2

Mason GJ 2010 Species differences in responses to captivity: stress, welfare and the comparative method. *Trends in Ecology & Evolution 25(12)*: 713-721. https://doi.org/10.1016/j.tree.2010.08.011

Mason GJ and Latham NR 2004 Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? *Animal Welfare 13*: S57-S69

Mason GJ and Veasey JS 2010 How should the psychological well-being of zoo elephants be objectively investigated? *Zoo Biology 29(2)*: 237-255. https://doi.org/10.1002/zoo.20256

McPhee ME and Carlstead K 2012 The importance of maintaining natural behaviors in captive mammals. In: Kleiman DG, Thompson KV and Kirk Baer C (eds) *Wild Mammals in Captivity: Principles and Techniques for Zoo Management, Second Edition*. University of Chicago Press: Chicago, USA

Meagher RK 2009 Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science 119(1-2)*: 1-14. https://doi.org/10.1016/j.applanim.2009.02.026

Meehan CL, Mench JA, Carlstead K and Hogan JN 2016 Determining connections between the daily lives of zoo elephants and their welfare: an epidemiological approach. *PLoS One 11(7)*: e0158124. https://doi.org/10.1371/journal.pone.0158124

Meller CL, Croney C and Shepherdson D 2007 Effects of rubberized flooring on Asian elephant behavior in captivity. *Zoo Biology 26(1)*: 51-61. https://doi.org/10.1002/zoo.20119

Menargues A, Urios V and Mauri M 2008 Welfare assessment of captive Asian elephants (*Elephas maximus*) and Indian rhinoceros (*Rhinoceros unicornis*) using salivary cortisol measurement. *Animal Welfare 17*: 305-312

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(7)*: e0155223. https://doi.org/10.1371/journal.pone.0155223

Millspaugh JJ, Burke T, Van Dyk G, Slotow R, Washburn BE and Woods RJ 2007 Stress response of working African elephants to transportation and safari adventures. *Journal of Wildlife Management 71(4)*: 1257-1260. https://doi.org/10.2193/2006-015

Millspaugh JJ and Washburn BE 2004 Use of fecal glucocorticoid metabolite measures in conservation biology research: considerations for application and interpretation. *General and Comparative Endocrinology 138*: 189-199. https://doi.org/10.1016/j.ygcen.2004.07.002

Morfeld KA, Meehan CL, Hogan JN and Brown JL 2016 Assessment of body condition in African (*Loxodonta africana*) and Asian (*Elephas maximus*) elephants in North American zoos and management practices associated with high body condition scores. *PLoS One 11(7)*: e0155146. https://doi.org/10.1371/journal.pone.0155146

Mormede P, Andanson S, Auperin B, Beerda B, Guemene D, Malmkvist J, Manteca X, Manteuffel G, Prunet P, van Reenen CG, Richard S and Veissier I 2007 Exploration of the hypothalamic-pituitary-adrenal function as a tool to evaluate animal welfare. *Physiology and Behaviour 92(3)*: 317-339. https://doi.org/10.1016/j.physbeh.2006.12.003

Nakagawa S and Cuthill IC 2007 Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews 82(4)*: 591-605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Palme R 2012 Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animals. *Animal Welfare 21(3)*: 331-337. https://doi.org/10.7120/09627286.21.3.331

PETA 2015 *Get Elephants Out of Zoos*. http://www.peta.org/issues/animals-in-entertainment/get-elephants-out-of-zoos.aspx

Poole J and Moss C 2008 Elephant sociality and complexity. In: Wemmer C and Christen CA (eds) *Elephants and Ethics: Toward a Morality of Coexistence*. Johns Hopkins University Press: Baltimore, MD, USA

Posta B, Huber R and Moore III DE 2013 The effects of housing on zoo elephant behavior: A quantitative case study of diurnal and seasonal variation. *International Journal of Comparative Psychology 26(1)*: 37-52

Ralph CR and Tilbrook AJ 2016 Invited review: The usefulness of measuring glucocorticoids for assessing animal welfare. *Journal of Animal Science 94(2)*: 457-470. https://doi.org/10.2527/jas.2015-9645

Ramanthan 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(2)*: 148-154. https://doi.org/10.1638/2007-0008R1.1

Rees PA 2004 Low environmental temperature causes an increase in stereotypic behaviour in captive Asian elephants (*Elephas maximus*) *Journal of Thermal Biology 29(1)*: 37-43. https://doi.org/10.1016/j.jtherbio.2003.10.004

Rees PA 2009 Activity budgets and the relationship between feeding and stereotypic behaviors in Asian elephants (*Elephas maximus*) in a Zoo. *Zoo Biology 28(2)*: 79-97. https://doi.org/10.1002/zoo.20200

Schmid J 1995 Keeping circus elephants temporarily in paddocks: the effects on their behaviour. *Animal Welfare 4*: 87-101

Schmid J, Heistermann M, Ganslober U and Hodges JK 2001 Introduction of foreign female Asian elephants (*Elephas maximus*) into an existing group: behavioural reactions and changes in cortisol levels. *Animal Welfare 10*: 357-372

Sharp T and Saunders G 2011 *A Model for Assessing the Relative Humaneness of Pest Animal Control Methods, Second Edition*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia

Sibbald B and Roland M 1998 Understanding controlled trials: Why are randomised controlled trials important? *BMJ 316(7126)*: 201. https://doi.org/10.1136/bmj.316.7126.201

Soltis J 2010 Vocal communication in African elephants (*Loxodonta africana*). *Zoo Biology 29(2)*: 192-209

Stoinski TS, Daniel E and Maple TL 2000 A preliminary study of the behavioral effects of feeding enrichment on African elephants. *Zoo Biology 19(6)*: 485-493. https://doi.org/10.1002/1098-2361(2000)19:6<485::AID-ZOO1>3.0.CO;2-5

Vanitha V, Thiyagesan K and Baskaran N 2011 Social life of captive Asian elephants (*Elephas maximus*) in Southern India: implications for elephant welfare. *Journal of Applied Animal Welfare Science 14(1)*: 42-58. https://doi.org/10.1080/10888705.2011.527603

Veasey J 2006 Concepts in the care and welfare of captive elephants. *International Zoo Yearbook 40(1)*: 63-79. https://doi.org/10.1111/j.1748-1090.2006.00063.x

Walter O 2010 *Management Guidelines for the Welfare of Zoo Animals: Elephants, Third Edition*. BIAZA: UK

Wells DL and Irwin RM 2008 Auditory stimulation as enrichment for zoo-housed Asian elephants (*Elephas maximus*). *Animal Welfare 17(4)*: 335-340

Wemelsfelder F and Mullan S 2014 Applying ethological and health indicators to practical animal welfare assessment. *Revue Scientifique et Technique de l'OIE 33(1)*: 111-120. https://doi.org/10.20506/rst.33.1.2259

Wemmer C, Krishnamurthy V, Shrestha S, Hayek LA, Thant M and Nanjappa KA 2006 Assessment of body condition in Asian elephants (*Elephas maximus*). *Zoo Biology 25(3)*: 187-200. https://doi.org/10.1002/zoo.20099

Whilde J and Marples N 2011 Effect of a birth on the behavior of a family group of Asian elephants (*Elephas maximus*) at Dublin Zoo. *Zoo Biology 31(4)*: 442-452. https://doi.org/10.1002/zoo.20408

https://doi.org/10.7120/09627286.27.3.235 Published online by Cambridge University Press

Whitham JC and Wielebnowski N 2009 Animal-based welfare monitoring: Using keeper ratings as an assessment tool. *Zoo Biology 28(6)*: 545-560. https://doi.org/10.1002/zoo.20281

Williams E, Bremner-Harrison S, Harvey N, Evison E and Yon L 2015 An investigation into resting behaviour in Asian elephants in UK zoos. *Journal of Zoo Biology 34*: 406-417. https://doi.org/10.1002/zoo.21235

Wilson M, Bloomsmith M and Maple T 2004 Stereotypic swaying and serum cortisol concentrations in three captive African elephants (*Loxodonta africana*). *Animal Welfare 13*: 39-43

Zoos' Forum Elephant Working Group 2010 *Elephants in UK Zoos: Zoos' Forum review of issues in elephant husbandry in UK zoos in light of the Report by Harris* et al *(2008)*. Wildlife Species Conservation Division, Defra, Bristol, UK