Animal Welfare 2021, 30: 1-18
ISSN 0962-7286
doi: 10.7120/09627286.30.1.001

*A critical review of animal-based welfare indicators for polar bears (*Ursus maritimus*) in zoos: Identification and evidence of validity*

CR Skovlund[*†‡], *MK Kirchner*[‡], *LW Moos*[‡], *N Alsted*[‡], *X Manteca*[§], *O Tallo-Parra*[§], *M Stelvig*[†] *and B Forkman*[‡]

[†] Research and Conservation, Copenhagen Zoo, Roskildevej 38, 2000 Frederiksberg, Denmark
[‡] Section of Animal Welfare and Disease Control, Department of Veterinary and Animal Sciences, University of Copenhagen, Groennegaardsvej 8, 1870 Frederiksberg C, Denmark
[§] Zoo Animal Welfare Education Centre (ZAWEC), Veterinary Faculty, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
* Contact for correspondence: ces@zoo.dk

Abstract

*Captive polar bears (*Ursus maritimus*) are well-documented as being prone to behavioural disorders and, as a result, their welfare is the cause of increasing concern. There is therefore a need for an evidence-based approach to the assessment of the welfare of this species and identification of valid welfare indicators is the first step towards achieving this. To this end, a critical evaluation of peer-reviewed literature was undertaken. Searches of Web of Science and Scopus took place in May 2020 for publications relevant to the welfare of captive polar bears which met inclusion criteria. Further, validity of extracted indicators was assessed via investigation of evidence of content, construct and criterion validity along with strength of evidence at publication-level. Database searches and snowballing unearthed 46 publications included for review. Identified indicators were sorted into nine behavioural, four physiological (based on physiological or biological sampling) and five physical (based on visual inspection) categories. Among behavioural indicators, the strongest evidence of validity was found for abnormal behaviour. For the physiological indicators, validity was only established for faecal glucocorticoid metabolite concentration. Content validity was assumed for all physical indicators. Generalisability and strength of evidence was generally compromised by low sample sizes and experimental limitations, and only a small number of papers investigated welfare indicators directly, resulting in a paucity of validated indicators. Potential welfare indicators that warrant further validation are highlighted. Overall, this review provides an overview of current valid and promising welfare indicators along with identified gaps in knowledge, relevant for the provision of a methodology for assessing and monitoring welfare of captive polar bears.*

Keywords: *animal welfare, behaviour, polar bear, welfare assessment, welfare indicators, zoo welfare*

Introduction

The role of modern zoos today extends beyond them being merely recreational facilities. Zoos are involved in education, research and, importantly, species conservation, yet the management of animal populations leads to inevitable trade-offs between the interests of the individual animal and broader species-specific interests (Cohen & Fennell 2016). In recent years, zoo animal welfare has become the subject of increasing focus from the media and NGOs and come under public scrutiny. The welfare of zoo-housed bears, in particular polar bears (*Ursus maritimus*), has been the subject of extra attention due to the species' well-documented susceptibility to behavioural abnormalities associated with captivity (Vickery & Mason 2003a). Stereotypic behaviour has been reported in between 55–80% of individuals (Mason *et al* 2007; Shepherdson *et al* 2013), occupying a large proportion of their waking hours (eg Brando *et al* 2018 reported a mean of 40% in various studies). Combined with evidence of poor reproductive success in the *ex situ* population (Curry *et al* 2015), concerns over the welfare of polar bears in zoos have emerged. Information on the current welfare status of the captive population is limited, however, due to a lack of validated indicators and standardised methods of assessment. Recently, Brando and colleagues (2018) reviewed current knowledge and the future direction for marine mammals in zoos and aquaria, while Maślak *et al* (2016) investigated the welfare of captive ursids in Poland. However, as yet, there have been no studies looking specifically into welfare states incorporating multiple variables for polar bears.

In itself, compromised zoo animal welfare constitutes an ethical concern due to humans' moral responsibility to animals. Additionally, poor welfare can lead to abnormal behaviour and reduced behavioural repertoire, weakening the educational value (Mason *et al* 2007; Rose *et al* 2017). Poor welfare may impact longevity and reproductive success undermining conservation efforts (Cameron & Ryan 2016). Considering polar bears' current 'Vulnerable' status on the

IUCN Red List (Wiig *et al* 2015) (with only 23,000 individuals left in the wild; Hamilton & Derocher 2019), the zoo community may well have an important role to play in conservation, promoting healthy *ex situ* populations. Indeed, in zoos, polar bears serve as a flagship species, advancing the public's awareness of the impact of climate change and conservation efforts. Striving for high standards of welfare is therefore a key responsibility of zoos for a variety of reasons, and research into how to assess, monitor and improve this welfare is imperative. The first step towards achieving this goal is identification of welfare indicators.

Evidence-based welfare assessment should be based on validated indicators and, as such, provide meaningful information on the welfare of the animal (Wemelsfelder & Mullan 2014). Key to this is the establishment of different levels of validity relative to the extent to which the connection of each indicator to animal welfare is projected. Williams and colleagues (2018) provide a practical framework for assessment of the different types of validity of welfare indicators for zoo animals, which we drew upon in this review. The terminology and validity levels used in the current paper, ranging from the weakest to the strongest are: content/face validity (whether there is a logical connection of the measure[s] to animal welfare), construct validity (whether the measure[s] assesses the broad area [construct] it is intended to measure, reflected in correlations between other related welfare measures) and criterion validity (the measure[s] is compared to an independent criterion measure or test [ie a gold standard] or the ability of these to, for example, differentiate between populations or environments of different levels of welfare) (for an in-depth description of the terms, see Williams *et al* 2018). Validity, on the publication level, is also important to consider when estimating indicator validity and, here, is evaluated based on a general synthesis of the included studies: this comprises both internal validity — relating to whether the experimental design is appropriate in light of the research in question and hence whether the experimental procedure is free from bias (Lehner 1979; O'Connor & Sargeant 2014) — and external validity, concerning whether the results from the sample population are generalisable to other contexts, populations or species (Lehner 1979). Additional important criteria are reliability (repeatability and observer agreement) and feasibility (time and cost efficiency) (Knierim & Winckler 2009).

This review incorporated the three main approaches to welfare: naturalness (Duncan & Petherick 1991), biological functioning (Broom 1986) and feelings (Dawkins 1998). This approach was chosen with the aim of achieving an holistic view of the animals' welfare state. To further facilitate welfare to be evaluated on a continuous scale ranging between poor and good (Hill & Broom 2009), multiple different indicators, such as behavioural, physiological and physical parameters pertinent both to negative and positive welfare, were included. This approach seeks to encompass societal and stakeholder opinion on the concept of animal welfare itself, thereby increasing the likelihood of acceptance of the resultant method of assessment. As such, the aim of this review was to: i) identify potential welfare indicators relevant to polar bears in zoos, which can be used to assess welfare on an individual level as well as monitor welfare over time and across conditions; and ii) assess validity and reliability of these indicators, along with strength of validity, and thus their potential to be included in evidence-based welfare assessments of polar bears.

Materials and methods

Search methods

Electronic database searches for relevant publications were carried out in May 2020 in Web of Science (WoS) (Core Collection) and Scopus. A search was conducted in each database (all years) for the words 'polar bear*' or '*Ursus maritimus*' in title, abstract and keywords (including Keywords Plus®). Also, a more specific search was conducted using different combinations of the keywords 'polar bear*', 'welfare', 'zoo*', 'captiv*' and 'animal welfare.' A broad search was conducted because more specialised searches using a combination of keywords gave rise to a limited number of articles. Snowballing was also performed to identify additional literature, and literature that was already in the authors' possession prior to searches was included as well. An initial review of both title and abstract took place to enable those articles not fulfilling the inclusion criteria to be discarded. Database searches, application of inclusion and exclusion criteria and extraction of details on publication-level (see below) were carried out by CRS.

Inclusion criteria

Publications that met all the following inclusion criteria were included in the review: (i) specific to polar bears in captivity (or including polar bears if reporting separate results on these); (ii) assessing welfare either directly or indirectly, including health, physical condition, physiology, behaviour, mental state or husbandry and management; (iii) including at least one animal-based welfare indicator; (iv) peer-reviewed (including conference proceedings); (v) available in English; and (vi) available in full.

Exclusion criteria

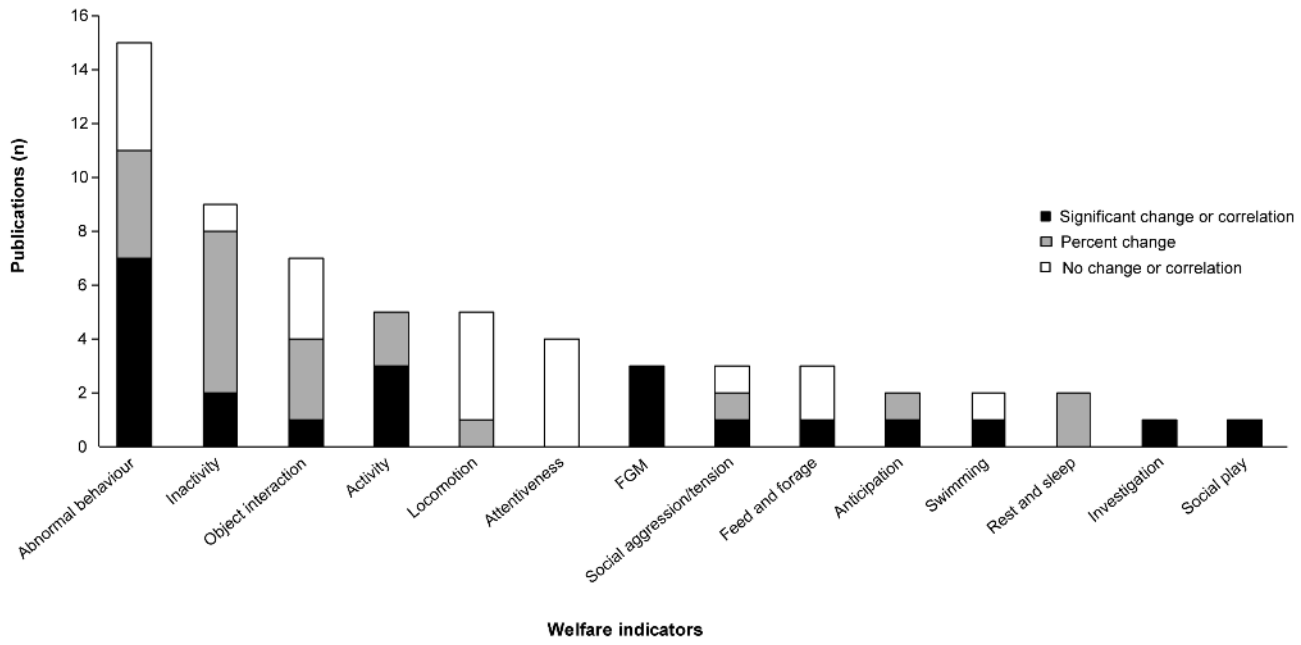
Publications not meeting the inclusion criteria were excluded from this review. This also included: (i) review articles; (ii) studies assessing population reproductive trends or morbidity rates that cannot be applied to the individual; and (iii) studies examining cub and maternal behaviour, health or nutrition. Although normal maternal (and cub) behaviour is relevant to welfare, an exclusion criterion was applied as these behaviours are likely to be specific to a particular life or developmental stage and therefore not suitable as general welfare indicators.

Validity of welfare indicators

The level of validity for each extracted indicator (content, construct and/or criterion) was based on the definition adapted by Williams *et al* (2018). Additionally, note was made whether the indicators varied in response to interventions, among groups or over different phases of the

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

Figure 1



An overview of the welfare indicators (behavioural and physiological) extracted from the critically reviewed publications on captive polar bears.

study or correlated with other animal-based indicators. It was also recorded whether significant results were provided or whether specific differences were reported in percentages. As such, construct and criterion validity were highlighted even when significant findings were not reported. Despite not being considered to rely on strong evidence, meaning a conclusion of true validity is not possible based on this, inclusion of change reported as percentage (Williams *et al* 2018) may provide useful information and was therefore noted in this review. Resource- and management-based parameters were also highlighted if shown to affect indicators, ie the impact (direction) on indicators (behaviour and physiology) and whether the effect or association was significant, was noted.

Critical review of publications — internal and external validity

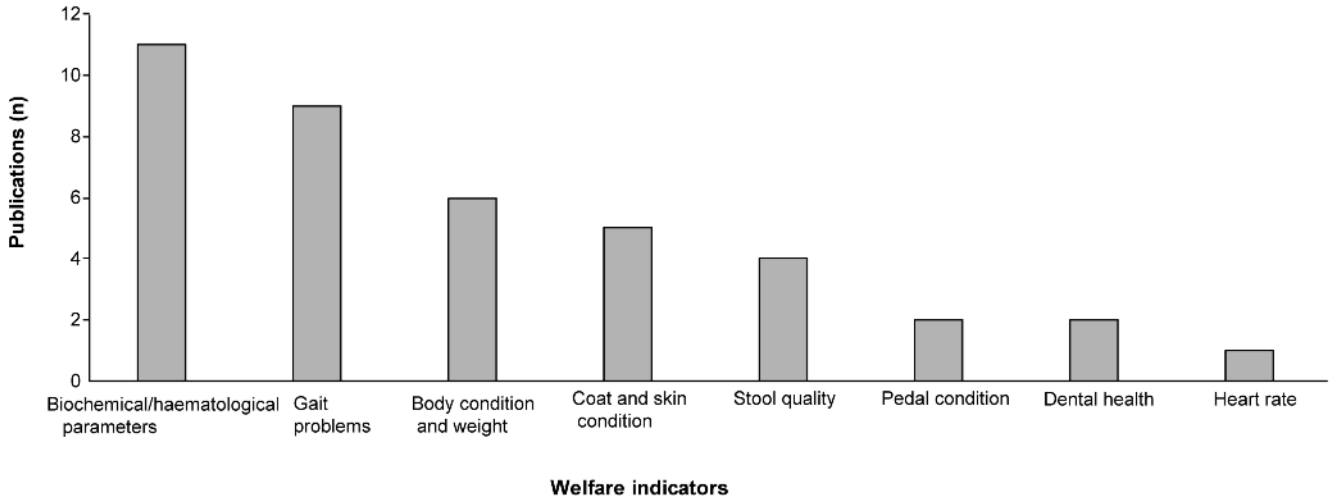
Publications meeting all the inclusion and exclusion criteria were critically reviewed to assess internal and external validity, thereby facilitating overall assessment of validity of each extracted indicator as regards its relevance to welfare. However, publications from which extracted indicators were identified related to disease or other health issues were not reviewed further, since content validity for these indicators was assumed. Internal validity was based on various parameters relating to strength of study design and whether or not appropriate steps were taken to reduce bias. The extraction process was inspired from the critical appraisal form laid out by Williams and colleagues (2018). Reliability was noted when it was reported as was the method of welfare assessment or any other objective of the study.

Results

Database searches yielded 3,796 results (2,518 in WoS and 1,278 in Scopus). Along with 'snowballing', this gave us 46 publications that reached inclusion criteria, and consequently were included for further review (for an overview of all included publications, see Table 1 in the supplementary material to papers published in *Animal Welfare*; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). The included publications were published between 1972 and 2020. Behavioural indicators were identified in 21 papers (Figure 1), physiological indicators in 16 papers (of which 12 were in relation to various disease and health problems), and physical indicators in 16 papers (all identified in health-related papers) (Figure 2). Various indicators, such as appetite and other symptoms of ill-health were identified in 14 papers.

Grouped into nine overall categories, behavioural indicators were identified as 'abnormal behaviour', 'activity', 'inactivity', 'rest and sleep', 'locomotion', 'feeding and foraging', 'environmental interaction', 'social behaviour' and 'other' which consisted of a group of indicators that did not fit into a category (anticipation and attentiveness). Identified physiological indicators were fewer and therefore not grouped: faecal glucocorticoid metabolite concentration (FGM), stool quality, heart rate and haematological and biochemical parameters. Overall categories of physical indicators included gait abnormalities, body condition (including weight), skin and coat condition, foot condition and dental health. Uncategorised indicators, such as reduced appetite, reduced activity and breathing abnormalities arising from impaired health were also highlighted. As mentioned, indicators

Figure 2



An overview of the welfare indicators (physical and physiological) extracted from the non-critically reviewed publications.

reported as symptoms of disease, malnutrition or other health issues were not reviewed further owing to presumed content validity since the indicators were reported as a consequence of or in conjunction with compromised health (see Figures 1 and 2 and, for an overview of unique indicators, the supplementary information in Table 1 (https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

Assessment of welfare or validation of indicators was rarely the focus of the studies. Instead, the various aims included: monitoring changes in behaviour in response to environmental interventions; investigating the effect of various environmental and management variables; investigating motivations and locomotor aspects of stereotypic behaviour; inspecting social dynamics; comparing behaviour or health parameters to *in situ* polar bears; change over time (seasons); and various health-related studies mostly consisting of disease cases (more information on each study can be found in Table 1; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). Studies with these various aims were nonetheless included because the findings may be utilised to assess overall strength of validity of indicators.

Sample size

The included publications were generally dominated by small sample sizes and only a relatively small number were multi-institutional. Sample size ranged from one to 55; median for the behavioural studies was two whereas the median for health-related studies was one. Number of included institutions ranged from one to 20 with a median of one for both publications on behaviour and health- and disease-related publications.

Behavioural indicators

Although assessment of validity of indicators was not the aim of most of the included papers, evidence of content, construct and criterion validity could be found for several of the identified indicators. Abnormal behaviour was identified in 15 publications and was the most frequently reported behaviour in the reviewed articles (Figure 1, Table 2; see also supplementary material for a full version of the table; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

Abnormal behaviour was reported as stereotypic behaviour, abnormal repetitive behaviour, terrestrial and aquatic route tracing, pacing, stutter stepping, head swinging, facial tics and coughing or huffing. Two studies found significant associations between pacing and an animal-based indicator; a positive association was found to FGM (Shepherdson *et al* 2013) and higher peak responses and variation was found in non-stereotyping individuals compared to stereotyping individuals (Shepherdson *et al* 2004). Moreover, in seven studies, abnormal behaviour was found to vary in response to interventions either to the environment or internal milieu: a reduction in stereotypic behaviour was observed after provision of 24-h access to an off-exhibit enclosure (Ross 2006), providing novel enrichment items (Canino & Powell 2010), providing naturalistic enrichment objects (Kutska 2009), feeding enrichment (Forthman *et al* 1992) and was terminated by manipulation of the serotonergic system by administration of fluoxetine (Poulsen *et al* 1996). One study found stereotypic behaviour increased in response to a novel odour (Linder *et al* 2020) while another found ambiguous results (Wechsler 1992). Pacing was quantifiably distinguished from non-repetitive locomotion by gait analysis and head height, along with differentiating intensity of pacing based on other variables in two papers (Cless *et al* 2015; Cless & Lukas 2017). Furthermore, various environmental factors were found to affect abnormal behaviour in four studies: stereotypic behaviour was positively associated with starve days (Ames 1993; Cremers & Geutjes 2012); traffic near exhibit and noise (> 70 dB) (Cremers & Geutjes 2012);

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

Table 2 Summary of the indicators identified in critically reviewed publications, and the respective association with other animal-based parameters and/or effect of interventions or influence of environmental factors.

Indicator	Identified in	Change or association in	Correlated or associated with (AB)	Affected by (RB and/or MB)	Evidence of validity
Abnormal behaviour	Ames (1993), Canino & Powell (2010), Cless & Lukas (2017), Cless *et al* (2015), Cremers & Geutjes (2012), Forthman *et al* (1992), Kelly *et al* (2014), Kutska (2009), Linder *et al* (2020), Poulsen *et al* (1996), Ross (2006), Shepherdson *et al* (2004, 2013), Wechsler (1991, 1992)	**Ames (1993), Cremers & Geutjes (2012), Kutska (2009), Linder *et al* (2020), Ross (2006), Shepherdson *et al* (2004, 2013)** Canino & Powell (2010), Forthman *et al* (1992), Kelly *et al* (2014), Poulsen *et al* (1996), Wechsler (1992)	**FGM+ Temperament: slow-to-approach+ (interest)–**	**Feeding– Naturalistic enrichment items (vs non-naturalistic items)– Odour marks -/+ Off-exhibit access– View out of exhibit– No of bears in group– Positive reinforcement training–** Novel enrichment items– Starve days+ Keeper presence– Noise (over 70dB)+ Feeding enrichment– Visitor density –/+ Drug administration (fluoxetine)–	Construct Criterion
Activity	Altman (1999), Ames (1993), Forthman *et al* (1992), Kelly *et al* (2014), Linder *et al* (2020)	**Altman (1999), Forthman *et al* (1992), Linder *et al* (2020)** Ames (1993), Kelly *et al* (2014)	None	**Enrichment objects+ Feeding enrichment+ Odour marks+**	Criterion
Anticipation	Ames (1993), Cless & Lukas (2017)	**Cless & Lukas (2017)** Ames (1993)	**Pacing intensity+**	**Pacing location (holding doors)+** Feeding strategy	Criterion
Attentiveness	Cremers & Geutjes (2012), Ross (2006), Wechsler (1991, 1992)	None	None	Off-exhibit access+	None
Feeding and foraging	Ames (1993), Folk *et al* (1973), Ross (2006)	**Ames (1993)**	Heart rate	**Feeding enrichment+**	Criterion
Faecal glucorticoid metabolite concentration	Hein *et al* (2020), Shepherdson *et al* (2004, 2013)	**Hein *et al* (2020), Shepherdson *et al* (2004, 2013)**	**Social tension+ Pacing + Temperament (interest)–**	**Animal transfer+ Environmental change+ Various disturbances+ Enclosure size (dry land)–**	Construct Criterion
Inactivity	Altman (1999), Ames (1993), Cremers & Geutjes (2012), Forthman *et al* (1992), Kelly *et al* (2014), Linder *et al* (2020), Poulsen *et al* (1996), Wechsler (1992)	**Forthman *et al* (1992), Linder *et al* (2020)** Altman (1999), Ames (1993), Cremers & Geutjes (2012), Kelly *et al* (2014), Poulsen *et al* (1996)	None	**Feeding enrichment –/0 Odour marks+** Object enrichment– Substrate+ Visitor density –/+ Drug administration (fluoxetine)+	Criterion
Locomotion	Cless *et al* (2015), Poulsen *et al* (1996), Ross (2006), Wechsler (1991, 1992)	Poulsen *et al* (1996)	None	Drug administration (fluoxetine)+	None
Object manipulation and play	Kutska (2009), Ames (1994), Altman (1999), Canino & Powell (2010), Ames (1993), Kuczaj *et al* (2002), Ross (2006)	**Kutska (2009)** Ames (1994), Altman (1999), Canino & Powell (2010)	None	**Naturalistic enrichment items (vs non-naturalistic items)+** Object enrichment+ Novel enrichment items+	Criterion
Rest and sleep	Canino & Powell (2010), Poulsen *et al* (1996)	Canino & Powell (2010) Poulsen *et al* (1996)	None	Object enrichment– Drug administration (fluoxetine)+	Criterion
Sniffing (investigation)	Wechsler (1992)	**Wechsler (1992)**	None	**Odour marks+**	Criterion
Social play	Ross (2006)	**Ross (2006)**	None	**Off-exhibit access+**	Criterion
Social aggression/tension	Ames (1993), Hein *et al* (2020), Renner & Kelly (2006)	**Hein *et al* (2020)** Ames 1993	**FGM+**	Mechanical feeder+	Construct
Swimming	Ross (2006), Wechsler (1991)	**Ross (2006)**	None	**Off-exhibit access+**	Criterion

Text in bold indicates significant findings (non-bold indicates % change not subjected to statistical analysis).
Effect/relationship with indicator is denoted by: '+': positive relationship; '–': negative relationship; '0': no observed change/effect in conjunction with '/' to indicate separate contradicting findings between individuals.
Abbreviations: AB: Animal-based; RB: Resource-based; MB: Management-based.

and was found negatively correlated with higher rates of enrichment, number of bears within the group, view out of exhibit and existence of an operant condition training programme (Shepherdson *et al* 2013). Since abnormal behaviour was associated with an animal-based indicator (FGM) and differed in various situations expected to impact welfare, evidence of both construct (convergent) (one paper) and criterion validity (six papers; one with significant change, percent change evident in three, and two reporting significant differences between components of non-repetitive locomotion and pacing) may be established (Tables 1 and 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

Inactivity, including immobility, passiveness, lying, sitting and standing, were identified in nine publications. Inactivity was seen to increase in response to provision of substrate (Cremers & Geutjes 2012), novel odour (Linder *et al* 2020) and administration of fluoxetine (Poulsen *et al* 1996), and decrease in response to object and (significantly) to feeding enrichment (one out of two individuals) (Forthman *et al* 1992; Altman 1999). Evidence of criterion validity was noted in four papers on account of observed change in response to provision of enriching conditions (significant in two papers and percent change in two). Rest and sleep were identified in two papers, for which sleep increased after fluoxetine administration (Poulsen *et al* 1996) and rest was reduced after provision of enrichment objects (Canino & Powell 2010) (criterion validity in one paper with percent change). Differences in overall activity levels were found between captive and wild bears (Ames 1993).

Environmental interaction covered manipulative and investigative behaviours (object manipulation, object play, digging and sniffing) and was investigated in eight publications. Sniffing (investigation) significantly increased when odour marks were introduced into the enclosure (Wechsler 1992) and provision of enrichment objects and introducing object novelty were shown to affect object interaction (Ames 1993; Altman 1999; Canino & Powell 2010), however significant results were only reported for provision of naturalistic compared to non-naturalistic items (Kutska 2009). Since the very nature of enrichment items is intrinsically enriching, evidence of criterion validity is thus assumed in two papers (percent change) for object manipulation, while investigation reached criterion validity in one paper (significant change).

Activity, including general levels of activity, was investigated in five publications. A significant increase in activity was seen in response to provision of enrichment objects (Altman 1999), odour object (Linder *et al* 2020) and feeding enrichment (Forthman *et al* 1992). Locomotion, including walking, running and swimming, was investigated in five publications. In one paper, swimming was seen to significantly increase after provision of off-exhibit access (Ross 2006). Non-stereotypic mobility was observed increasing in response to fluoxetine treatment (Poulsen *et al* 1996). As a result of changes in response to enriching alterations to the environment, criterion validity is evident for activity in four papers (significant change in three and percent change in one) and for locomotion (swimming) in one paper (significant change).

Social behaviours (four papers) were identified as social play, aggressive behaviour and social tension, other non-aggressive interactions and inter-individual distance. Social tension was significantly associated with FGM (Hein *et al* 2020), and social play was observed to significantly increase when 24-h access to the off-exhibit area was given (Ross 2006), thereby demonstrating evidence of both construct (one paper) and criterion validity (one paper). Percent change was observed for aggressive interactions across feeding conditions in one paper (Ames 1993), however this was most likely a result of competitive behaviour.

Feeding and foraging were investigated in three papers; foraging levels were found to increase significantly after provision of feeding enrichment (one paper; Ames 1993) which may indicate criterion validity. Further, one paper demonstrated a slowing of the heart rate during feeding (Folk *et al* 1973).

Other behaviours included anticipation (two papers) which was mentioned as both a behavioural measure and in relation to qualitative rating of individuals as pacing anticipatorily. Cless and Lukas (2017) found significantly more intensive pacing (assessed via gait analysis) in bears that had been independently classified as anticipatory pacers compared to other individuals; criterion validity may be established based on this (significant findings in one paper and percent change in another).

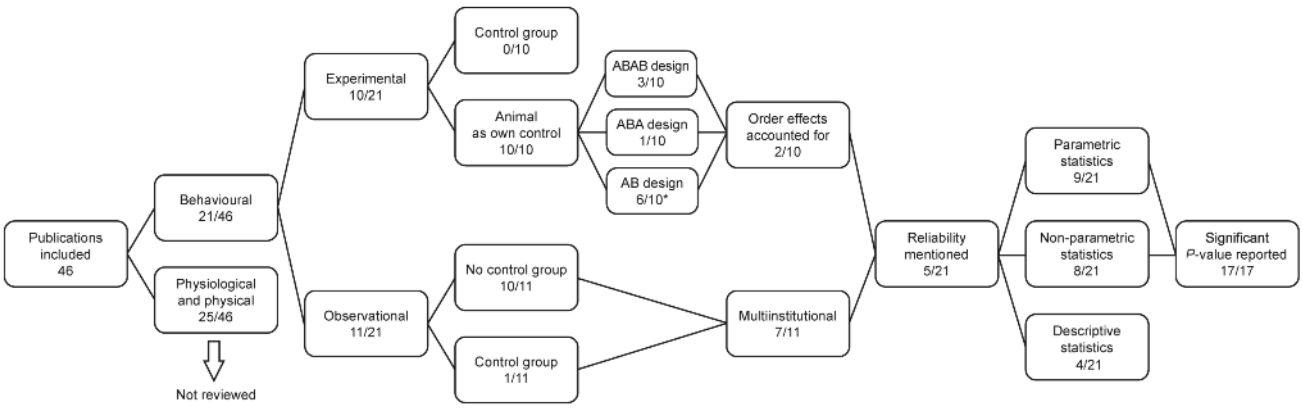
Attentiveness (mentioned as general attentiveness, ie 'looking up' or 'attention to staff') was identified in four publications. Only a small increase in frequency was reported after provision of off-exhibit access (Ross 2006).

Physiological indicators

Physiological indicators were less frequently identified in the reviewed literature compared to behavioural indicators. FGM (three publications) were significantly positively correlated with pacing (Shepherdson *et al* 2013) and social tension (Hein *et al* 2020), and higher peak levels and variation were found in non-stereotyping compared to stereotyping individuals (Shepherdson *et al* 2004). Furthermore, FGM was positively associated with various disturbances and environmental change (Hein *et al* 2020), and animal transport (Shepherdson *et al* 2013; Hein *et al* 2020). Construct and criterion validity were found for FGM (two papers with significant associations and one with significant change). Eleven studies found several biochemical, haematological and other abnormalities in relation to various diseases, whereby two found sub-optimal vitamin levels in serum (vitamin D) associated with antebrachial fractures, and lower vitamin A and E levels in captive versus free-ranging individuals (Kenny *et al* 1998; Lin *et al* 2005). In one study, heart rate was found to lower during feeding (Folk *et al* 1973) and stool quality (including observations of diarrhoea) was extracted from four papers (Figure 2 and Table 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

Figure 3



Overview of highlighted features related to strength and utility of study designs for validating welfare indicators. Each box indicates the number of publications which utilised each given feature (indicated left of '/') out of the total number of reviewed publications within the category (indicated right of '/'). Indicators found to be significant with parametric analysis were object manipulation, activity, stereotypic behaviour, passiveness, faecal glucocorticoid metabolite concentration and social tension. Indicators found to be significant with non-parametric analysis were stereotypic behaviour, pacing intensity, foraging, activity, inactivity, social play, swimming, sniffing, locomotion, rest/sleep and anticipation. Indicators explained with descriptive statistics were heart rate, object interaction, stereotypic behaviour, inactivity, activity and aggressive interactions. ABA and ABAB: Repeated treatments design; AB: Repeated measures design; A: Baseline; B: Experimental (Swaisgood & Shepherdson 2005). * AB design assumed when not possible to discern.

Physical indicators

All the pin-pointed physical indicators were found in publications describing various health problems and related medical symptoms. Gait abnormalities were referred to in nine publications, in the form of various levels of lameness, including limb weakness, fractures and flexor rigidity. Other commonly identified physical indicators included body condition and weight (including muscle wastage) (six papers), skin and coat condition (including urine scalding) (five papers), dental health (tooth fractures and gum condition) (two papers) and pedal condition (pedal abscesses) (two papers) (Figure 2 and Table 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

Other ungrouped indicators reported in relation to ill-health included, amongst others, reduced appetite (ten papers) and acute symptoms, such as vomiting and regurgitation (five papers), abnormal breathing (two papers), heavy salivation (one paper) and polyuria-polydipsia (one paper) (Table 2; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material), however these were not further reviewed.

Overall strength of evidence in relation to welfare indicator validity

Although few of the included studies actually set out to validate welfare indicators or assess welfare, strength of validation in this review is based on adopted study features at publication-level. Details of study features related to strength of overall indicator validation of the included publications are thus highlighted in Figure 3. Overall, few of the reviewed papers adopted a broad range of study features which increase robustness of findings and hence internal validity, almost certainly due to the practical limitations that accompany conducting research in a zoo setting. Ten studies were experimental in nature and included a form of intervention thought to affect welfare in a semi-controlled manner. Eleven papers adopted an observational study design, of which two undertook an epidemiological approach, including large sample sizes across multiple institutions, and seven were multi-institutional. External validity could generally be considered at moderate to high levels since the results were assumed to be generalisable inter-individually and inter-institutionally, due to the exclusive inclusion of captive polar bears in the review and since the included studies were carried out primarily in zoos (Figure 3).

Discussion

Study and indicator validity

The aim of this review was to identify valid animal-based welfare indicators for use in the assessment of captive polar bear welfare. A review of the literature enabled several potential indicators to be successfully identified. Validity was based on type of validity reached (ie content, construct or criterion validity) along with an evaluation of internal and external validity assessed on publication level. The latter was used in support of an overall assessment of each indicator and to identify shortcomings in evidence. The extraction and validation process was inspired by the appraisal tool proposed by Williams and colleagues (2018), which is particularly useful for literature on sparsely researched species, providing a suitable approach to establish type of validity to meet the objective of this study. Criterion validity was reached for several of the behavioural indicators, while construct validity was reached solely for pacing and social tension. Construct and criterion validity was found for the physiological indicator, FGM and content validity was established for the physical indicators. Thus, a degree of validity was attained for most of the

identified indicators. The aim of the included studies was less to do with validating animal-based welfare indicators, and more about investigating changes in behaviour in response to various interventions or observing the effect of environmental variables on behaviour and, occasionally, the physiological response. This can provide a valuable insight into potential resource- and management-based indicators to be used as a proxy when no validated animal-based indicators are available. The lack of validation studies results in a scarcity of thoroughly validated animal-based welfare indicators for polar bears, however, useful information could still be extracted and several potential welfare indicators were identified.

By adopting the approach of Williams *et al* (2018), criterion validity was established in several publications. Criterion validity was reached because several of the studies compared different environments or introduced short-term welfare interventions (eg comparison of enriched and non-enriched environments, or the introduction of a novel enrichment strategy). Such changes or introductions were assumed, in this review, to have an effect on welfare, and hence, by the adopted definition, the reported behavioural changes following this would indicate criterion validity. This approach is more loosely defined compared to other definitions (eg Meagher 2009; Belshaw *et al* 2015) since it does not merely rely upon a 'gold standard'. This led to the inclusion of studies using interventions that should be interpreted with caution concerning the establishment of validity, eg fluoxetine treatment (changes internal state and behaviour) or provision of enrichment objects (potential problem with interpretation concerning cause and effect) but, for the purposes of this review, these were still included. However, this definition also allowed additional publications to be included which is helpful when extracting and reviewing welfare indicators in such a seldom researched species as the polar bear. The change in behavioural indicators observed across different conditions may indicate their relationship with welfare (ie negative or positive) and may reflect the corresponding change in welfare states, that is imposed by these conditions, of the observed individuals. These indicators could thus be interpreted as risk signals, however not as truly validated measures (Normando *et al* 2018). This is useful for monitoring welfare over time and across conditions, where altered behaviour may reflect a change in welfare state. Since behaviour is part of an organisms' defence mechanism, allowing it to cope with the environment (Palme 2012), behavioural changes serve as an important indicator of alterations to welfare status, that are both practical and non-invasive (Watters *et al* 2019). Further scrutiny of trends across studies showing, for example, similar responses elicited in similar conditions and variables, may still provide useful information about indicators' welfare loading (see Table 1 and Table 2 in supplementary material; https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

As previously stated, validity was highlighted even when based purely on reported percent change and non-significant findings. Applying more stringent criteria would have led to information potentially being neglected. This review therefore serves to not only identify valid indicators but also provide an overview of emerging trends regarding valence of indicators.

Studies that focused on impaired health did not undergo further review as they were assumed to be valid (content validity) due to their association with impaired health. Indeed, an individual experiencing health-related issues as a result of illness or ageing, may still encounter more positive than negative episodes (Mellor 2016) and, thus, be considered to exist within the positive spectrum of welfare. Nonetheless, these indicators are included in this review since they may indicate or cause pain or discomfort and be interpreted within the range of negative to neutral valence. Additionally, indicators identified in this review, including weight, body condition, appetite, pedal, skin, coat and dental conditions are common to geriatric polar bears and are thus important to include in health inspections (AZA Bear TAG 2009), not to mention holistic welfare assessments.

The remaining publications had various aims and related experimental set-ups, which made it difficult to compare internal validity and validation of individual indicators. However, we have sought to extract information on study-level to assess internal and external validity. The included studies were generally dominated by small sample sizes which, in certain circumstances, hindered parametric statistics or the reporting of statistical significance, where descriptive statistics were described instead. Since this review relied upon significant findings to establish true criterion or construct validity, this impeded the extraction of valid indicators. However, even experimental studies with small sample sizes can take steps to limit bias, ensuring robustness and increasing internal validity, for example, by employing randomisation, repeated treatments, withdrawal, multiple baseline designs (Saudargas & Drummer 1996; Todman & Dugard 2001) and randomisation. Although experimental studies have the ability to control for unwanted variation and are thus considered by some to be superior in this regard, behavioural observational studies may also adopt features to reduce bias, such as 'blocking' and 'matching' (Dawkins 2007), as well as making use of control groups, blinding (to avoid expectation bias) (Tuyttens *et al* 2016) and inter-observer reliability, similar to experimental studies, thereby ensuring internal validity (Dawkins 2007). Due to the impracticalities associated with carrying out behavioural research in a zoo setting, only a limited number of the studies took additional measures to increase internal validity or reported doing so, thereby hindering critical appraisal of the literature. Research here is often complicated by the limited availability of study subjects and difficulty standardising settings. Also, experimental research is further complicated by the risk of compromising welfare via exposure to differing treatment regimes, such as withdrawing enrichment. Since many studies were based on individual or a few bears and failed to report actions designed to limit bias, overall indicator validity needs to be interpreted cautiously. However, by compiling results across studies, tendencies on indicator-level may still be evident. The multi-institution studies naturally reach higher levels of external validity (Saudargas

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

& Drummer 1996), however, external validity was compromised in many of the studies that included only one or a few institutions. Although institutions commonly differ in such factors as social housing, management and resources and geographical location, findings may still indicate what ought to be observed when replicating studies in other institutions. Two studies employed an epidemiological approach whereby sample size was greatly increased and by which variation can instead be used as an advantage to strengthen findings. Although such a design requires ample resources, these studies provide a valuable insight into the association between animal-based measures and variables, and results are considered to be generalisable.

Behavioural indicators of welfare

Abnormal behaviour (mainly stereotypies, such as pacing and other types of route tracing) was the primary focus of several of the reviewed studies. Consequently, it was the most frequently identified indicator with most evidence, reaching both construct and criterion validity; it correlated with other behavioural and physiological indicators that it is theoretically related to, and decreased in response to interventions likely introducing a change in welfare. Moreover, at the publication-level, it was observed as having an inverse relationship with indicators assumingly reflecting positive welfare (eg play and object interaction). Taking trends across studies into account, a reduction in abnormal behaviour was observed in conjunction with increased social play (Ross 2006), increased swimming (Ross 2006), increased object play (Canino & Powell 2010), increased object manipulation (Kutska 2009), reduced rest (Canino & Powell 2010), increased immobility and non-stereotypic mobility (Poulsen *et al* 1996). Although this is demonstrated across various situations, it may indicate divergent validity, as a reduction in abnormal behaviour had a tendency to be followed by an increase in indicators of what may be considered positive welfare. Similar results have been found in other ursids, where abnormal behaviour has been reported as decreasing in response to various forms of enrichment (eg Carlstead *et al* 1991; Forthman *et al* 1992; Maślak *et al* 2016; Wagman *et al* 2018), improved dental condition (Maslak *et al* 2013), commonly followed by an increase in positive behaviours, such as play (Koene 1998) and foraging (Maslak *et al* 2013), or increased during situation-related frustration (Waroff *et al* 2017), underscoring stereotypies as reflecting reduced welfare. The general consensus is that abnormal behaviour is a reflection of impaired welfare, and may be a consequence of, for example, increased stress arising in response to a chronically aversive environment (Garner 2005) or an underlying medical problem (Krebs *et al* 2018). Furthermore, associations have been shown with, amongst others, reduced behavioural diversity (Miller *et al* 2016) and behavioural flexibility (Vickery & Mason 2003b), impaired reproduction (Carlstead & Shepherdson 1994; Carlstead *et al* 1999) and elevated glucocorticoid concentration (GC) (Wielebnowski *et al* 2002). Interpretation in relation to welfare, however, is not necessarily straightforward since stereotypic behaviour may be a symptom of neurological dysfunction (Cabib 2006), be a relic from previous sub-optimal captive conditions or be serving as an individual's coping strategy to relieve symptoms of these conditions (Mason & Latham 2004). However, the consideration of changes in stereotypic behaviour over time (Williams *et al* 2018) remains an effective welfare assessment indicator since it tends to change in response to improvement or deterioration in conditions in the expected direction of valence. Compelling evidence of abnormal behaviour does not necessarily equate to it being the most significant, feasible or reliable indicator, or that it should be used as a primary indicator. As a result of previously mentioned barriers to interpretation, it is vital that stereotypies are assessed in combination with other indicators of welfare, health status and life history of the animal. Abnormal behaviour was also found to vary over time and season (Ames 1993; Kelly *et al* 2014), perhaps reflecting internal motivational states and changes in management (or lack of corresponding changes). However, inter-individual patterns exist (Ames 1993; Kelly *et al* 2014) and studies have reported different trends of abnormal and other behavioural patterns across seasons. Individuality, seasonality and reproductive phase should thus be considered in any assessment of welfare over time. Overall, abnormal behaviour was observed decreasing in response to what may be considered positive changes and conditions (except in the case of introduction of dog odour; Linder *et al* 2020), and increasing in response to deteriorating circumstances. So, the role abnormal behaviour plays in signifying compromised welfare may cautiously be inferred from this and may be suitable as a parameter to detect welfare changes over time and across conditions. Location of pacing may also provide a valuable insight into the underlying motivation of the behaviour (Cless & Lukas 2017).

Although interrelated, in this review, activity and inactivity level were treated separately since they were reported as separate measures in the studies. Both reached criterion validity as the indicators changed in scenarios in which welfare was presumably altered and, in many cases, reduced activity was reported as a sign of impaired health. Although inactivity (eg awake inactivity) is functionally separate from true rest, it was often impossible to differentiate between them in the papers and the two were therefore merged in this review. Changes in activity levels were observed across conditions that seemingly give rise to both improved and reduced welfare, thus failing to provide a straightforward indication of valence. That said, excessive activity or inactivity can, in some instances, indicate compromised welfare. Captive polar bears are often deprived of competition, with abundant resources and, in contrast to their wild conspecifics, have no pressing need to preserve energy through high levels of inactivity. So, in some cases, excess inactivity may be a result of a captive environment failing to facilitate species-appropriate behaviour, giving rise to concerns regarding apathy (Manteca *et al* 2016), depression-like states, boredom, chronic stress or illness (Fureix & Meagher 2015), and may therefore constitute a viable indicator of negative affect. However, in other contexts, inactivity may indicate positive

low-arousal affective states (ie inactivity as a result of the provision of soft substrate; Cremers & Geutjes 2012). *In situ* polar bears have been reported to spend a large proportion of their time inactive (42 and 87%; Stirling 1974; Ames 1993, respectively) with behaviour influenced by season, ie climatic conditions and food availability (Messier *et al* 1992). *Ex situ* polar bears are reported as being more active than their wild counterparts (Ames 1993). Taking natural behaviour as a reference point, high inactivity levels need not necessarily indicate compromised welfare in polar bears. However, captivity offers a radically different milieu to the Arctic, and so excess inactivity, reference levels, and assumed welfare implications warrant further attention. It is also crucial to be able to tell different forms of inactivity apart (eg by defining awake inactivity; Harvey *et al* 2019) prior to drawing any conclusions about affective state. Many of the reviewed studies were not seeking to investigate activity levels, therefore, this differentiation may be further complicated by inconsistency in definition of terms. Although no 'optimal' or 'cut-off' level of activity/inactivity is currently in place in terms of polar bear welfare, monitoring activity levels over time may still help identify sudden changes, eg as a result of impaired health, yet it is imperative that activity levels are considered contextually. Locomotion was another frequently identified indicator yet change (in swimming) was only reported in one study in response to 24-h off-exhibit access, which also increased play and reduced stereotypy. Locomotion (walking) has been found to correlate with enriching environmental variables in elephants (*Loxodonta africana* and *Elephas maximus*) (Holdgate *et al* 2016) and may thus indirectly reflect engagement with the environment and be a sign of positive affect, but more research is needed to validate this indicator. Lastly, when incorporating activity and/or locomotion into welfare assessment, it is important to take into account the individual in question since age, sex and season may have exerted an influence, particularly where polar bears are concerned. For example, more rest was observed in older brown bears (*Ursus arctos*) (Montaudouin & Le Pape 2004) and walking distance was negatively correlated with age in elephants (Holdgate *et al* 2016); this may not in itself necessarily constitute a welfare concern, although increasing age may be accompanied by factors impinging on welfare, such as impaired health.

Environmental interaction, often referred to as object manipulation, was another commonly identified group of indicators. It reached criterion validity due to an increase in enriching conditions, although there are a lack of significant findings. Object interaction was found to increase in conjunction with reduced stereotypies and rest, and when activity and foraging increased, perhaps underscoring its potential role as a welfare indicator. Object manipulation was often identified in studies investigating the effectiveness or preference of enrichment items rather than validation as a welfare indicator, and it was not always possible to assess whether it was related to extrinsic or inquisitive exploration or play. Object manipulation may be placed within the category of behaviours relating to competence-building agency, and may thus indicate positive welfare since, besides facilitating the expression of natural behaviours, it enhances motor and cognitive skills as well as the gathering of knowledge (Špinka 2019). This behaviour is thus assumed not only to be pleasurable, but to also improve future welfare due to the capacity building allowing optimised ability to cope with future challenges (Špinka 2019). Sniffing was also extracted and may be the only indicator to truly reflect inquisitive exploration, although it was not investigated in relation to welfare. In other ursids, variable-time feeding enrichment schedules have promoted exploratory behaviours and concurrently reduced abnormal behaviours (Wagman *et al* 2018). Although little evidence can be extracted from this review, object interaction has the potential to be included in welfare assessments as a sign of positive affect considering the general link to welfare, although further validation is needed; this is especially true for exploration and object play. Irrespective of this, environmental interaction may be used as an indication of the effectiveness of enrichment strategy.

As a result of the appetitive aspect of extrinsic exploration, foraging levels may be closely related to this behaviour. Feeding and foraging were not however frequently identified in the reviewed literature. Although foraging met criteria for criterion validity as it increased in response to feeding enrichment, it should be considered cautiously owing to the broad assumption taken in this review, ie whereby imposed enrichment is considered an enriched condition. An appropriate feeding strategy may nonetheless stimulate foraging and exploration and may thereafter enhance environmental engagement. *In situ* polar bears spend between 35–50% of their time hunting (Stirling 1998) and their foraging patterns are highly adapted to an environment with resources scattered spatially, necessitating great distances to be traversed before feeding sites are reached. Further, they spend several hours still hunting at breathing holes. As such, finding ways to mimic these activities in captivity and stimulating foraging levels to take up a greater proportion of bears' time can be extremely challenging. Levels of foraging and food manipulation may thus be a reflection of feeding strategy efficiency, in terms of promoting species-specific behaviours and enrichment level. Previously, a number of feeding enrichment schemes have effectively increased food manipulation and exploration and reduced abnormal behaviours and passivity in various ursids (Forthman *et al* 1992; McGowan *et al* 2010; Wagman *et al* 2018). For polar bears, however, there is little evidence of these functioning as valid indicators, even though a reduced appetite was often reported as being a symptom of impaired health (Alroy *et al* 1980; Morris *et al* 1989; LaDouceur *et al* 2014; Morrison *et al* 2017) and should be monitored closely (despite not being well-validated indicators at the present time).

Although social behaviours were identified less frequently, evidence of both criterion and construct validity were found for social play and social tension, respectively, owing to it increasing in response to an assumed improvement in management (24-h off-exhibit access), and being correlated

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

with FGM. Bears were found to actively maintain and increase inter-individual distance (Renner & Kelly 2006), yet group size was negatively correlated with pacing, which could indicate a positive effect of social housing (Shepherdson *et al* 2013). One study found concurrent bear activity to influence stereotypic behaviour, whereby stereotypy increased in one bear when conspecifics were out of sight, increased in all bears when conspecifics were inactive and all bears showed less stereotypy when conspecifics were active (Kelly *et al* 2014), perhaps due to increased social vigilance. The relationship between pacing and group size as regards welfare, may not therefore be straightforward. Evidence exists suggesting that appropriate social housing may enhance welfare even in solitary species, eg as observed in Southern tamandua (*Tamandua tetradactyla*) (Catapani *et al* 2019). Social housing may give rise to social buffering (Kikusui *et al* 2006) and enable affiliative behaviours, such as allogrooming, maternal care and sexual gratification, all of which have the potential to confer positive emotions (Mellor 2015). Captive polar bears are often housed with conspecifics despite being considered a solitary species (except during mating, cub-rearing and occasional male aggregations) (Derocher & Stirling 1990). Since female polar bears in the wild avoid males outside of the breeding season to protect offspring (Demaster & Stirling 1981), being housed in close proximity to the opposite sex may increase stress levels, relative to time of year and reproductive phase. This complicates social housing of polar bears, as appropriateness of social housing may be reliant not only on the individuals in question (eg relatedness in brown bears; Montaudouin & Le Pape 2005, personality type; Tetley & O'Hara 2012 and sex) but also on the time of year. Social housing may give rise to overt displacement or inability to reach resources as a result of competition (Ames 1993), and social instability can be a source of chronic stress in captivity (Morgan & Tromborg 2007). Regular assessment of social dynamics is therefore imperative. That said, social interaction can be rare in polar bears and alternative indicators may need to be identified and validated. In other zoo-housed species, aggressive encounters have been associated with increased GC, eg in dorcas gazelles (*Gazella dorcas*) (Salas *et al* 2016) and chimpanzees (*Pan troglodytes*) (Yamanashi *et al* 2018), and reduced GC have been associated with higher levels of affiliation and play in rhesus monkeys (*Macaca mulatta*) (Wooddell *et al* 2017), which provide some evidence of validity for these indicators in other species. Social behaviour may thus comprise both positive and negative valence welfare indicators.

There was evidence of criterion validity of anticipatory behaviour, or rather anticipation, as anticipatory rated individuals performed higher intensity pacing than non-anticipatory rated individuals. Intensity of anticipation may indicate increased reward sensitivity and thus reduced welfare (Watters 2014). Moreover, it is proposed that stereotypic behaviour, especially in bears, often stems from frustrated appetitive behaviour (Montaudouin & Le Pape 2004) which could thus be thwarted, feeding-related anticipation. Conversely, anticipatory behaviour may also be interpreted with positive valence, where short and frequent bouts of anticipatory behaviour may be followed by positive emotions in response to anticipated events — the frequency and intensity may be used to distinguish the valence of this behaviour (Watters 2014). Discriminating between anticipatory behaviour and other repetitive behaviour (in the case of different underlying causes) may be cumbersome, however, for example, location (eg close to holding doors), head height and direction, alertness and intensity may aid interpretation (Swaisgood *et al* 2001; Cless *et al* 2015; Cless & Lukas 2017). More evidence is needed to validate this measure for polar bears and to investigate the underlying motivations for both anticipation and various forms of stereotypic behaviour.

Attentiveness was only identified in a few publications and, although not further validated as an indicator in itself, the finding that polar bears remained attentive while stereotyping may indicate this to be an anticipation-related behaviour. Consequently, this may be used to assess whether a behaviour is, in fact, anticipatory or non-anticipatory repetitive behaviour. Inattentiveness may also be a sign of disengagement and apathy and may thus be inferred as a negative valence indicator. Attentiveness has been utilised in Qualitative Behaviour Assessment for zoo-housed elephants (Yon *et al* 2019) and could therefore serve as a sensible qualitative indicator. Both anticipatory behaviour and attentiveness may be relevant to keeper and/or routine dependence, possibly constituting a welfare issue and, for that reason, may be important to be incorporated into polar bears' welfare assessment.

Temperament was mentioned in one publication and despite not intrinsically being a welfare indicator, the traits bold and shy were associated with less and more pacing and FGM, respectively (Shepherdson *et al* 2013). Thus, an ability to cope in captivity may depend on certain traits, rendering personality relevant to individual welfare. Through understanding how different personalities respond to various environmental factors in captivity, at-risk individuals may be identified before problems arise and management tailored accordingly (Shepherdson *et al* 2013; Richter & Hintze 2019).

Based on the evidence that has emerged from this review, strongest evidence of validity currently exists for stereotypic behaviour, indicating compromised welfare. The findings pertaining to the remaining indicators would benefit from further research; seeking to validate and test for reliability. Validated indicators of positive affect are generally lacking across species.

Physiological indicators of welfare

FGM was not a commonly identified indicator yet reached construct (convergent) and criterion validity as it was positively correlated with pacing and social tension, and highly elevated after transportation, respectively. On the other hand, lower peak corticoid levels and variation was also evident in stereotyping compared to non-stereotyping individuals, implying that the latter react more strongly to stressors or find the environment more stressful (Shepherdson *et al* 2004). A link between high GC levels and abnormal behaviour has been found in other captive species, eg clouded leopards (*Neofelis nebulosi*)

(Wielebnowski *et al* 2002), and has been associated with lower levels of play and exploration in tufted capuchins (*Cebus paella*) (Byrne & Suomi 2002), highlighting the potential link between GC and welfare. However, the stress response in different welfare states or conditions is not always straightforward; hypothalamic-pituitary-adrenal axis activity has been reported to increase, decrease or remain constant in response to chronic stressors in various species (Pawluski *et al* 2017) further complicating interpretation of this indicator. It is well-established that severe or prolonged stressors can inhibit an organism's ability to restore homeostasis, thereby adversely impacting welfare in terms of behaviour, immunocompetence, health and reproduction (Moberg 2000; Möstl & Palme 2002). However, interpretation is further complicated by the fact that the neuroendocrine system also becomes activated during instances beneficial to the individual (eg mating), and the stress response is affected by individual factors (eg physiological stage, age and life history; Mormède *et al* 2007; Williams *et al* 2018). Moreover, in the case of chronic stressors, the organism may cease to elicit an appropriate endocrine response altogether, further complicating the matter (Mormède *et al* 2007). For example, Pawluski and colleagues (2017) found low cortisol levels associated with depressive-like behaviour in domestic horses (*Equus caballus*). Although several caveats need to be considered, assessing GC in conjunction with other indicators (behavioural, physical and preferably physiological) may still provide valuable information about the welfare of captive animals (Williams *et al* 2018; Palme 2019). In stark contrast to plasma GC (which represents short time-frames of the adrenocortical response), FGM is a far more stable measure, reflecting the average plasma GC over several hours due to the excretion lag time (Hein *et al* 2020). Faecal collection further mitigates the risk of the sampling method (eg blood sampling) affecting the stress response and is useful for mid- to long-term monitoring of the stress response plus assessing how environmental change can affect GC (Palme 2012; Hein *et al* 2020). Importantly, retention and peak excretion time can impact results (Hein *et al* 2020). For example, an average gastrointestinal transit time of seven hours (Hein *et al* 2020) and peak excretion time after three days (Shepherdson *et al* 2013) have been reported for captive polar bears, which illustrates why daily and/or long-term sampling is beneficial. Although faecal sampling was the only medium found to assess GC in captive polar bears in this review, others may be considered, such as serum, saliva or hair. For example, serum cortisol concentrations were compared for captive grizzly bears (*Ursus arctos horribilis*) trained for blood collection and chemically immobilised individuals, where serum cortisol levels were only evident in the latter group, likely due to the stressful nature of immobilisation (Joyce-Zuniga *et al* 2016). Hair cortisol concentrations (HCC) have been successfully utilised for polar bears *in situ* (Bechshøft *et al* 2011, 2012; Macbeth *et al* 2012; Mislan *et al* 2016) and may provide a feasible indicator for long-term, retrospective stressors and welfare. HCC have also been found suitable for long-term retrospective assessments of GC in Asiatic black bears (*Ursus thibetanus*), however seasonal excretion patterns have been reported which need to be taken into account (Malcolm *et al* 2013). However to the authors' knowledge, season has had no reported impact on FGM in captive polar bears (Bryant & Roth 2018; Hein *et al* 2020).

The remaining studies on physiological indicators investigated biochemical or haematological abnormalities as well as stool quality in relation to impaired health. Sub-optimal vitamin levels of *ex situ* bears were found, and fractures were associated with subnormal vitamin D levels (an outcome, incidentally, found to occur with greater frequency than previously thought). This highlights the need for adequate monitoring and to ensure proper nutrition and health for captive polar bears, since feeding (and thus nutrition) differs for captive animals compared to wild conspecifics. When measuring physiological parameters, samples that can be taken non-invasively (eg stool, hair or saliva), are preferable as they do not compromise current welfare. Monitoring stool condition can provide an insight into nutrition and/or potential disease, and a rating scale has been devised for this purpose (Lintzenich *et al* 2006). Although a proportion of the captive polar bear population is found in geographical areas where the ambient temperature exceeds that of the Arctic, thermal stress was not identified as an indicator in this review. Methods utilised to assess body heat in polar bears include infra-red imagery (Øritsland *et al* 1974; Smith *et al* 2020) and thermal stress has the potential to be assessed via heat shock proteins (Park *et al* 2015) or respiration rate, although no validated indicator feasible for a zoo setting is currently available. Research into thermal stress in captive polar bears and animal-based indicators to aid assessment is therefore needed.

Physical indicators of welfare

Several physical indicators were identified, however solely in papers investigating impaired health. None of the papers assessed validation of indicators and consequently no indicators were correlated with other indicators or changed across conditions (except in the case of improvement in health). Nonetheless, as impaired health can adversely impact welfare (eg in the case of chronic pain compromising affective state by inducing chronic stress; Fureix & Meagher 2015), it is included in this review. Gait abnormalities were the most common indicator mentioned in various forms, and reported as a result of fractures, lumbar stenosis and myasthenia gravis, amongst others. Pedal abscesses were also reported, and is a recognised condition in polar bears that requires monitoring (AZA Bear TAG 2009). Assessment of pedal health and development of a gait score system should therefore enhance the welfare of captive polar bears. Degenerative joint disease is reported as being prevalent in ageing ursids, however instead of manifesting as lameness, symptoms often include reduced activity; behavioural inspection is thus needed to ensure proper identification (Bourne *et al* 2010). Body condition was the second most identified physical indicator and

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

found to diminish in response to various health issues. Maintaining an adequate body mass is crucial in the prevention and management of lameness (both of which are inter-related) (Bourne *et al* 2010). A fatness index has been developed and validated for wild polar bears (Stirling *et al* 2008) which may be utilised to monitor fluctuations in body condition in captive bears. Muscle mass is important to consider in conjunction with fatness levels, since muscle wastage may occur as a consequence of disease or restricted or barren enclosures, limiting mobility. Coat and skin condition were also identified and should be included in the health assessment. For example, alopecia has been reported as one of the more common conditions in captive polar bears (AZA Bear TAG 2009), occurring also in wild individuals (Bowen *et al* 2015). It may be caused by bacterial, fungal or parasitic infections (Bourne *et al* 2010), seasonal allergies, trauma, water quality and reproductive hormonal imbalances (AZA Bear TAG 2009), potentially causing discomfort and thermal stress, and requiring monitoring. Dental issues were also reported. Oral inspection is important owing to the prevalence of broken teeth and other oral health concerns in captive carnivores, eg caused by an inappropriate diet or oral stereotypies in ursids (Wenker *et al* 1999). Old age and inadequate nutrition seem commonly reported causes of ill health in captive polar bears, and although some of the included papers pre-date significant improvements in polar bear management, health issues may still be prevalent and pass undetected at times. As an example, end-stage renal disease was found to be the most diagnosed cause of death or euthanasia in captive polar bears in the United States in a relatively recent study (LaDouceur *et al* 2014). Regular health assessments are especially important when we consider that the captive bear population is becoming ever more geriatric (Kitchener 2004; Bourne *et al* 2010). Additionally, several behavioural indicators, such as reluctance to enter water, excess scratching, inappetence and increased inactivity, were reported in cases of morbidity (eg Smith & Cordes 1972; Lacasse & Gamble 2006; Eo & Kwon 2014), and are therefore important to include alongside health indicators. Reviewing health-related indicators was not the main aim of this review but broadening the search method to include various types of literature would likely result in additional reports and more information on these indicators.

Animal welfare implications

Addressing animal welfare scientifically has important outcomes for its improvement. Welfare assessment should be carried out based on evidence, and this review sought to facilitate this by evaluating the validity of the indicators found for zoo-housed polar bears. This was done by searching the literature for validated and feasible indicators of welfare, which may be sampled directly from the animal. Traditionally, institutional approaches to zoo animal welfare have relied upon husbandry recommendations and resource and management variables. These recommendations can come in the form of Association of Zoos and Aquariums (AZA) animal care manuals, which are comprised of knowledge about the biological and physical needs of species (eg the *Polar Bear Care Manual*; AZA Bear TAG 2009). Although valuable resources in terms of appropriate husbandry standards to uphold, these manuals provide little information on the actual experiences of the individuals being housed (Whitham & Wielebnowski 2009) and prove less useful for assessment and monitoring of welfare at the individual animal level. To construct an holistic method of animal welfare assessment, multiple variables including output, and input measures (when necessary), should be compiled into a welfare assessment protocol that strives to incorporate all aspects of welfare. Protocols are currently widely implemented for livestock and, of late, are progressing into the realms of captive wildlife (Hill & Broom 2009; Rees 2015). So far, institutional welfare protocols have been developed, or partly developed, for elephants (Yon *et al* 2019), dorcas gazelle (Salas *et al* 2018), bottlenose dolphins (*Tursiops truncatus*) (Clegg *et al* 2015) and pygmy blue tongue skinks (*Tiliqua adelaidensis*) (Benn *et al* 2019), yet no protocol currently exists for zoo-housed polar bears. This review initiates the first step toward development of such a protocol, which will facilitate current assessment of welfare at the individual-level, monitoring changes in welfare over time as well as identification of factors affecting welfare in captivity and, consequently, promoting the welfare and conservation of these animals.

Conclusion

This review identified several behavioural, physiological and physical parameters from peer-reviewed literature that have the potential to serve as welfare indicators for zoo-housed polar bears. Based on the evidence that has emerged from this review, evidence of validity currently exists for abnormal behaviour, reflecting negative welfare. Promising behavioural indicators warranting further research are environmental interaction (including direct manipulation and investigative behaviour), activity/inactivity, feeding and foraging, play, social behaviours, anticipatory behaviour and attentiveness. FGM may serve as a useful indicator deployed in conjunction with other indicators. Indicators that were not critically reviewed yet assumed to be relevant to welfare, included stool quality, body condition and weight, gait abnormalities, pedal health, dental health and coat and skin condition. Besides body condition and stool quality, development of scoring schemes for these are recommended along with routine assessments. Research and subsequent evidence of validity of welfare indicators is currently sparse and studies with the objective of validating indicators for polar bears and other ursids are therefore necessary.

Acknowledgements

This research was funded by Innovation Fund Denmark. We are grateful to Dr Lydia Kolter for providing additional publications to review.

Conflict of interest

Two of the authors (CRS) and (MS) are employed at Copenhagen Zoo.

References

Alroy J, Baldwin D and Maschgan ER 1980 Multiple beta cell neoplasms in a polar bear: An immunohistochemical study. *Veterinary Pathology 17*: 331-337. https://doi.org/10.1177/0300985880017003 07

Altman JD 1999 Effects of inedible, manipulable objects on captive bears. *Journal of Applied Animal Welfare Science 2*: 123-132. https://doi.org/10.1207/s15327604jaws0202_4

Ames A 1993 The behaviour of captive polar bears. *UFAW Animal Welfare Research Report No 5* pp 67. UFAW: Wheathampstead, UK

AZA Bear TAG 2009 *Polar Bear (*Ursus maritimus*) Care Manual*. Association of Zoos and Aquariums, Silver Spring, MD, USA

Bechshøft T, Rigét FF, Sonne C, Letcher RJ, Muir DCG, Novak MA, Henchey E, Meyer JS, Eulaers I, Jaspers VLB, Eens M, Covaci A and Dietz R 2012 Measuring environmental stress in East Greenland polar bears, 1892-1927 and 1988-2009: What does hair cortisol tell us? *Environment International 45*: 15-21. https://doi.org/10.1016/j.envint.2012.04.005

Bechshøft T, Sonne C, Dietz R, Born EW, Novak MA, Henchey E and Meyer JS 2011 Cortisol levels in hair of East Greenland polar bears. *Science of the Total Environment 409*: 831-834. https://doi.org/10.1016/j.scitotenv.2010.10.047

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

Benn AL, McLelland DJ and Whittaker AL 2019 A review of welfare assessment methods in reptiles, and preliminary application of the Welfare Quality® protocol to the pygmy blue-tongue skink, *Tiliqua adelaidensis*, using animal-based measures. *Animals 9*: 1-22. https://doi.org/10.3390/ani9010027

Bourne DC, Cracknell JM and Bacon HJ 2010 Veterinary issues related to bears (*Ursidae*). *International Zoo Yearbook 44*: 16-32. https://doi.org/10.1111/j.1748-1090.2009.00097.x

Bowen L, Keith Miles A, Stott J, Waters S and Atwood T 2015 Enhanced biological processes associated with alopecia in polar bears (*Ursus maritimus*). *Science of the Total Environment 529*: 114-120. https://doi.org/10.1016/j.scitotenv.2015.05.039

Brando S, Broom DM, Acasuso-Rivero C and Clark F 2018 Optimal marine mammal welfare under human care: Current efforts and future directions. *Behavioural Processes 156*: 16-36. https://doi.org/10.1016/j.beproc.2017.09.011

Broom DM 1986 Indicators of poor welfare. *British Veterinary Journal 142*: 524-526. https://doi.org/10.1016/0007-1935(86)90109-0

Bryant JL and Roth TL 2018 Annual faecal glucocorticoid metabolite concentrations in pregnant and pseudopregnant polar bears (*Ursus maritimus*) in North American zoos. *Journal of Zoo and Aquarium Research 6*: 6-11

Byrne G and Suomi SJ 2002 Cortisol reactivity and its relation to home-cage behavior and personality ratings in tufted capuchin (*Cebus apella*) juveniles from birth to six years of age. *Psychoneuroendocrinology 27*: 139-154. https://doi.org/10.1016/S0306-4530(01)00041-5

Cabib S 2006 The neurobiology of stereotypy II: The role of stress. In: Mason G and Rushen J (eds) *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare Welfare, Second Edition* pp 227-255. CABI: Wallingford, Oxford, UK. https://doi.org/10.1079/ 9780851990040.0227

Cameron EZ and Ryan SJ 2016 Welfare at multiple scales: Importance of zoo elephant population welfare in a world of declining wild populations. *PLoS One 11*: e0158701. https://doi.org/10.1371/journal.pone.0158701

Canino W and Powell D 2010 Formal behavioral evaluation of enrichment programs on a zookeeper's schedule: A case study with a polar bear (*Ursus maritimus*) at the Bronx Zoo. *Zoo Biology 29*: 503-508. https://doi.org/10.1002/zoo.20247

Carlstead K, Fraser J, Bennett C and Kleiman DG 1999 Black rhinoceros (*Diceros bicornis*) in US zoos: II. Behavior, breeding success, and mortality in relation to housing facilities. *Zoo Biology 18*: 35-52. https://doi.org/10.1002/(SICI)1098-2361(1999)18:1<35::AID-ZOO5>3.0.CO;2-L

Carlstead K, Seidensticker J and Baldwin R 1991 Environmental enrichment for zoo bears. *Zoo Biology 10*: 3-16. https://doi.org/10.1002/zoo.1430100103

Carlstead K and Shepherdson D 1994 Effects of environmental enrichment on reproduction. *Zoo Biology 13*: 447-458. https://doi.org/10.1002/zoo.1430130507

Catapani ML, Pires JSR and Vasconcellos AS 2019 Single- or pair-housed: Which is better for captive southern tamanduas? *Journal of Applied Animal Welfare Science 22*: 289-297. https://doi.org/10.1080/10888705.2018.1508352

Clegg ILK, Borger-Turner JL and Eskelinen HC 2015 C-Well: The development of a welfare assessment index for captive bottlenose dolphins (*Tursiops truncatus*). *Animal Welfare 24*: 267-282. https://doi.org/10.7120/09627286.24.3.267

Cless IT and Lukas KE 2017 Variables affecting the manifestation of and intensity of pacing behavior: A preliminary case study in zoo-housed polar bears. *Zoo Biology 36*: 307-315. https://doi.org/10.1002/zoo.21379

Cless IT, Voss-Hoynes HA, Ritzmann RE and Lukas KE 2015 Defining pacing quantitatively: A comparison of gait characteristics between pacing and non-repetitive locomotion in zoo-housed polar bears. *Applied Animal Behaviour Science 169*: 78-85. https://doi.org/10.1016/j.applanim.2015.04.002

Cohen E and Fennell D 2016 The elimination of Marius, the giraffe: Humanitarian act or callous management decision? *Tourism Recreation Research 41*: 168-176. https://doi.org/10.1080/02508281.2016.1147211

Cremers PW and Geutjes SL 2012 The cause of stereotypic behaviour in a male polar bear (*Ursus maritimus*). In: Spink AJ, Grieco F, Krips OE, Loijens LWS and Noldus PHZ (eds) *Proceedings of Measuring Behavior* pp 338-340. 28-31 August 2012, Utrecht, The Netherlands

Curry E, Safayi S, Meyerson R and Roth TL 2015 Reproductive trends of captive polar bears in North American zoos: A historical analysis. *Journal of Zoo and Aquarium Research 3*: 99-106

Dawkins MS 1998 Evolution and animal welfare. *The Quarterly Review of Biology 73*: 305-328. https://doi.org/10.1086/420307

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

Dawkins MS 2007 *Observing Animal Behaviour: Design and Analysis of Quantitative Data*. Oxford University Press: Oxford, UK

Demaster BDP and Stirling I 1981 *Ursus maritimus*. *Mammalian Species 145*: 1-7. https://doi.org/10.2307/3503828

Derocher AE and Stirling I 1990 Observations of aggregating behaviour in adult male polar bears (*Ursus maritimus*). *Canadian Journal of Zoology 68*: 1390-1394. https://doi.org/10.1139/z90-207

Duncan IJ and Petherick JC 1991 The implications of cognitive processes for animal welfare. *Journal of Animal Science 69*: 5017-5022. https://doi.org/10.2527/1991.69125017x

Eo KY and Kwon OD 2014 Dermatitis caused by *Dermatophilus congolensis* in a zoo polar bear (*Ursus maritimus*). *Pakistan Veterinary Journal 34*: 560-562

Folk GE, Berberich JJ and Sanders DK 1973 Bradycardia of the polar bear. *Arctic 26*: 78-79. https://doi.org/10.14430 /arctic2900

Forthman DL, Elder SD, Bakeman R, Kurkowski TW, Noble CC and Winslow SW 1992 Effects of feeding enrichment on behavior of three species of captive bears. *Zoo Biology 11*: 187-195. https://doi.org/10.1002/zoo.1430110307

Fureix C and Meagher RK 2015 What can inactivity (in its various forms) reveal about affective states in non-human animals? A review. *Applied Animal Behaviour Science 171*: 8-24. https://doi.org/10.1016/j.applanim.2015.08.036

Garner JP 2005 Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes. *ILAR Journal 46*: 106-117. https://doi.org/10.1093/ilar.46.2.106

Hamilton SG and Derocher AE 2019 Assessment of global polar bear abundance and vulnerability. *Animal Conservation 22*: 83-95. https://doi.org/10.1111/acv.12439

Harvey ND, Moesta A, Kappel S, Wongsaengchan C, Harris H, Craigon PJ and Fureix C 2019 Could greater time spent displaying waking inactivity in the home environment be a marker for a depression-like state in the domestic dog? *Animals 9*: 1-19. https://doi.org/10.3390/ani9070420

Hein A, Palme R, Baumgartner K, von Fersen L, Woelfing B, Greenwood AD, Bechshoft T and Siebert U 2020 Faecal glucocorticoid metabolites as a measure of adrenocortical activity in polar bears (*Ursus maritimus*). *Conservation Physiology 8*: 1-16. https://doi.org/10.1093/conphys/coaa012

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

Holdgate MR, Meehan CL, Hogan JN, Miller LJ, Soltis J, Andrews J and Shepherdson DJ 2016 Walking behavior of zoo elephants: Associations between GPS-measured daily walking distances and environmental factors, social factors, and welfare indicators. *PLoS One 11*: 1-17. https://doi.org/10.1371 /journal.pone.0150331

Joyce-Zuniga NM, Newberry RC, Robbins CT, Ware JV, Jansen HT and Nelson OL 2016 Positive reinforcement training for blood collection in grizzly bears (*Ursus arctos horribilis*) results in undetectable elevations in serum cortisol levels: A preliminary investigation. *Journal of Applied Animal Welfare Science 19*: 210-215. https://doi.org/10.1080/10888705.2015.1126523

Kelly KR, Harrison ML, Size DD and MacDonald SE 2014 Individual effects of seasonal changes, visitor density, and concurrent bear behavior on stereotypical behaviors in captive polar bears (*Ursus maritimus*). *Journal of Applied Animal Welfare Science 18*: 1-15. https://doi.org/10.1080/10888705.2014.924832

Kenny DE, Irlbeck NA, Chen TC, Lu Z and Holick MF 1998 Determination of vitamins D, A, and E in sera and vitamin D in milk from captive and free-ranging polar bears (*Ursus maritimus*), and 7-dehydrocholesterol levels in skin from captive polar bears. *Zoo Biology 17*: 285-293. https://doi.org/10.1002/(SICI)1098-2361(1998) 17:4<285::AID-ZOO3>3.0.CO;2-5

Kikusui T, Winslow JT and Mori Y 2006 Social buffering: Relief from stress and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences 361*: 2215-2228. https://doi.org /10.1098/rstb.2006.1941

Kitchener AC 2004 The problems of old bears in zoos. *International Zoo News 51*: 282-293

Knierim U and Winckler C 2009 On-farm welfare assessment in cattle: Validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare 18*: 451-458

Koene P 1998 Adaptation of blind brown bears to a new environment and its residents: Stereotypy and play as welfare indicators. *Ursus 10*: 579-587

Krebs BL, Marrin D, Phelps A, Krol L and Watters JV 2018 Managing aged animals in zoos to promote positive welfare: A review and future directions. *Animals 8*: 1-22. https://doi.org/ 10.3390/ani8070116

Kuczaj S, Thad Lacinak OF and Trone M 2002 Keeping environmental enrichment enriching. *International Journal of Comparative Psychology 15*: 127-137

Kutska D 2009 Variation in visitor perceptions of a polar bear enclosure based on the presence of natural vs unnatural enrichment items. *Zoo Biology 28*: 292-306. https://doi.org/ 10.1002/zoo.20226

Lacasse C and Gamble KC 2006 Tracheitis associated with *Bordetella bronchiseptica* in a polar bear (*Ursus maritimus*). *Journal of Zoo and Wildlife Medicine 37*: 190-192. https://doi.org/10.1638/05-055.1

LaDouceur EEB, Garner MM, Davis B and Tseng F 2014 A Retrospective study of end-stage renal disease in captive polar bears (*Ursus maritimus*). *Journal of Zoo and Wildlife Medicine 45*: 69-77. https://doi.org/10.1638/2013-0071R.1

Lehner PN 1979 *Handbook of Ethological Methods*. Garland STPM Press: New York, USA

Lin RC, Engeli E, Prowten AW, Erb HN, Ducharme NG and Goodrich LR 2005 Antebrachial fractures in four captive polar bears (*Ursus maritimus*). *Veterinary Surgery 34*: 358-365. https://doi.org/10.1111/j.1532-950X.2005.00055.x

Linder AC, Gottschalk A, Lyhne H, Langbak MG, Jensen TH and Pertoldi C 2020 Using behavioral instability to investigate behavioral reaction norms in captive animals: Theoretical implications and future perspectives. *Symmetry 12*: 603. https://doi.org/10.3390/sym12040603

Lintzenich BA, Ward AM, Edwards MS, Griffin ME and Robbins CT 2006 Polar bear nutrition guidelines. *Polar Bears International, AZA Bear TAG* pp 65. https://nagonline.net/wp-content/uploads/2018/11/Lintzenich-et-al.-2006-Polar-Bear-Nutrition-Guidelines.pdf

Macbeth BJ, Cattet MRL, Obbard ME, Middel K and Janz DM 2012 Evaluation of hair cortisol concentration as a biomarker of long-term stress in free-ranging polar bears. *Wildlife Society Bulletin 36*: 747-758. https://doi.org/10.1002/wsb.219

Malcolm KD, McShea WJ, Van Deelen TR, Bacon HJ, Liu F, Putman S, Zhu X and Brown JL 2013 Analyses of fecal and hair glucocorticoids to evaluate short- and long-term stress and recovery of Asiatic black bears (*Ursus thibetanus*) removed from bile farms in China. *General and Comparative Endocrinology 185*: 97-106. https://doi.org/10.1016/j.ygcen.2013.01.014

Manteca X, Amat M, Salas M and Temple D 2016 Animal-based indicators to assess welfare in zoo animals. *CAB Reviews 11*: 1-10. https://doi.org/10.1079/PAVSNNR201611010

Maślak R, Sergiel A, Bowles D and Paśko Ł 2016 The welfare of bears in zoos: A case study of Poland. *Journal of Applied Animal Welfare Science 19*: 24-36. https://doi.org/10.1080/10888705.2015.1071671

Maślak R, Sergiel A and Hill SP 2013 Some aspects of locomotory stereotypies in spectacled bears (*Tremarctos ornatus*) and changes in behavior after relocation and dental treatment. *Journal of Veterinary Behavior: Clinical Applications and Research 8*: 335-341. https://doi.org/10.1016/j.jveb.2013.05.004

Mason G, Clubb R, Latham N and Vickery S 2007 Why and how should we use environmental enrichment to tackle stereotypic behaviour? *Applied Animal Behaviour Science 102*: 163-188. https://doi.org/10.1016/j.applanim.2006.05.041

Mason GJ and Latham NR 2004 Can't stop, won't stop: Is stereotypy a reliable animal welfare indicator? *Animal Welfare 13*: 57-69

McGowan RTS, Robbins CT, Alldredge JR and Newberry RC 2010 Contra-freeloading in grizzly bears: Implications for captive foraging enrichment. *Zoo Biology 29*: 484-502. https://doi.org/10.1002/zoo.20282

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

Mellor DJ 2015 Positive animal welfare states and encouraging environment-focused and animal-to-animal interactive behaviours. *New Zealand Veterinary Journal 63*: 9-16. https://doi.org/10.1080/00480169.2014.926800

Mellor DJ 2016 Updating animal welfare thinking: Moving beyond the 'Five Freedoms' towards 'A life worth living.' *Animals 6*: 21. https://doi.org/10.3390/ani6030021

Messier F, Taylor MK and Ramsay MA 1992 Seasonal activity patterns of female polar bears (*Ursus maritimus*) in the Canadian Arctic as revealed by satellite telemetry. *Journal of Zoology 226*: 219-229. https://doi.org/10.1111/j.1469-7998.1992.tb03835.x

Miller LJ, Pisacane CB and Vicino GA 2016 Relationship between behavioural diversity and faecal glucocorticoid metabolites: A case study with cheetahs (*Acinonyx jubatus*). *Animal Welfare 25*: 325-329. https://doi.org/10.7120/09627286.25.3.325

Mislan P, Derocher AE, St Louis VL, Richardson E, Lunn NJ and Janz DM 2016 Assessing stress in Western Hudson Bay polar bears using hair cortisol concentration as a biomarker. *Ecological Indicators 71*: 47-54. https://doi.org/10.1016/j.ecolind.2016.06.034

Moberg GP 2000 Biological response to stress: Implications for animal welfare. In: Moberg GP and Mench JA (eds) *The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare* pp 1-12. CAB International: Wallingford, UK. https://doi.org/10.1079 /9780851993591.0000

Montaudouin S and Le Pape G 2004 Comparison of the behaviour of European brown bears (*Ursus arctos arctos*) in six different parks, with particular attention to stereotypies. *Behavioural Processes 67*: 235-244. https://doi.org/10.1016/j.beproc.2004.02.008

Montaudouin S and Le Pape G 2005 Comparison between 28 zoological parks: Stereotypic and social behaviours of captive brown bears (*Ursus arctos*). *Applied Animal Behaviour Science 92*: 129-141. https://doi.org/10.1016/j.applanim.2004.10.015

Morgan KN and Tromborg CT 2007 Sources of stress in captivity. *Applied Animal Behaviour Science 102*: 262-302. https://doi.org/10.1016/j.applanim.2006.05.032

Mormède P, Andanson S, Aupérin B, Beerda B, Guémené 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 Behavior 92*: 317-339. https://doi.org/10.1016/j.physbeh.2006.12.003

Morris PJ, Legendre AM, Bowersock TL, Brooks DE, Krahwinkel DJ, Shires GMH and Walker MA 1989 Diagnosis and treatment of systemic blastomycosis in a polar bear (*Ursus maritimus*) with itraconazole. *Journal of Zoo and Wildlife Medicine 20*: 336-345

Morrison JF, Vakharia K and Moreland DB 2017 Lumbar laminectomy in a captive, adult polar bear (*Ursus maritimus*). *Surgical Neurology International 8*: 1-7. https://doi.org/10.4103/sni.sni_133_17

Möstl E and Palme R 2002 Hormones as indicators of stress. *Domestic Animal Endocrinology 23*: 67-74. https://doi.org/10.1016/S0739-7240(02)00146-7

Normando S, Pollastri I, Florio D, Ferrante L, Macchi E, Isaja V and de Mori B 2018 Assessing animal welfare in animal-visitor interactions in zoos and other facilities. A pilot study involving giraffes. *Animals 8*: 153. https://doi.org/10.3390/ani8090153

O'Connor AM and Sargeant JM 2014 Critical appraisal of studies using laboratory animal models. *ILAR Journal 55*: 405-417. https://doi.org/10.1093/ilar/ilu038

Øritsland NA, Lentfer JW and Ronald K 1974 Radiative surface temperatures of the polar bear. *Journal of Mammalogy 55*: 459-461. https://doi.org/10.2307/1379018

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

Palme R 2019 Non-invasive measurement of glucocorticoids: Advances and problems. *Physiology and Behavior 199*: 229-243. https://doi.org/10.1016/j.physbeh.2018.11.021

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press

Park KA, Duncan C, Noi P, Sonne C, Laidre K, Obbard M, Wiig Ø, Aars J, Regehr E, Gustafson LL and Atwood T 2015 Establishing a definition of polar bear (*Ursus maritimus*) health: A guide to research and management activities. *Science of the Total Environment Journal 514*: 371-378. https://doi.org/10.1016/j.scitotenv.2015.02.007

Pawluski J, Jego P, Henry S, Bruchet A, Palme R, Coste C and Hausberger M 2017 Low plasma cortisol and fecal cortisol metabolite measures as indicators of compromised welfare in domestic horses (*Equus caballus*). *PLoS One 12*: 1-18. https://doi.org/10.1371/journal.pone.0182257

Poulsen EM, Honeyman V, Valentine PA and Teskey GC 1996 Use of fluoxetine for the treatment of stereotypical pacing behavior in a captive polar bear. *Journal of the American Veterinary Medical Association 209*: 1470-1474

Rees PA 2015 *Studying Captive Animals: A Workbook of Methods in Behaviour, Welfare and Ecology*. Wiley-Blackwell: New York, USA

Renner MJ and Kelly AL 2006 Behavioral decisions for managing social distance and aggression in captive polar bears (*Ursus maritimus*). *Journal of Applied Animal Welfare Science 9*: 233-239. https://doi.org/10.1207/s15327604jaws0903_5

Richter SH and Hintze S 2019 From the individual to the population – and back again? Emphasising the role of the individual in animal welfare science. *Applied Animal Behaviour Science 212*: 1-8. https://doi.org/10.1016/j.applanim.2018.12.012

Rose PE, Nash SM and Riley LM 2017 To pace or not to pace? A review of what abnormal repetitive behavior tells us about zoo animal management. *Journal of Veterinary Behavior: Clinical Applications and Research 20*: 11-21. https://doi.org/10.1016/j.jveb.2017.02.007

Ross SR 2006 Issues of choice and control in the behaviour of a pair of captive polar bears (*Ursus maritimus*). *Behavioural Processes 73*: 117-120. https://doi.org/10.1016/j.beproc.2006.04.003

Salas M, Manteca X, Abáigar T, Delclaux M, Enseñat C, Martínez-Nevado E, Quevedo MÁ and Fernández-Bellon H 2018 Using farm animal welfare protocols as a base to assess the welfare of wild animals in captivity — case study: Dorcas gazelles (*Gazella dorcas*). *Animals 8*: 1-14. https://doi.org/10.3390/ani8070111

Salas M, Temple D, Abáigar T, Cuadrado M, Delclaux M, Enseñat C, Almagro V, Martínez-Nevado E, Quevedo MÁ, Carbajal A, Tallo-Parra O, Sabés-Alsina M, Amat M, Lopez-Bejar M, Fernández-Bellon H and Manteca X 2016 Aggressive behavior and hair cortisol levels in captive Dorcas gazelles (*Gazella dorcas*) as animal-based welfare indicators. *Zoo Biology 35*: 467-473. https://doi.org/10.1002/zoo.21323

Saudargas RA and Drummer LC 1996 Single subject (small N) research designs and zoo research. *Zoo Biology 15*: 173-181. https://doi.org/10.1002/(SICI)1098-2361(1996)15:2<173::AID-ZOO7>3.0.CO;2-8

Shepherdson D, Carlstead K and Wielebnowski N 2004 Cross-institutional assessment of stress responses in zoo animals using longitudinal monitoring of faecal corticoids and behaviour. *Animal Welfare 13*: 105-113

Shepherdson D, Lewis KD, Carlstead K, Bauman J and Perrin N 2013 Individual and environmental factors associated with stereotypic behavior and fecal glucocorticoid metabolite levels in zoo housed polar bears. *Applied Animal Behaviour Science 147*: 268-277. https://doi.org/10.1016/j.applanim.2013.01.001

Smith CF and Cordes DO 1972 Dermatitis caused by *Dermatophilus congolensis* infection in polar bears (*Thalactos maritimus*). *The British Veterinary Journal 128*: 366-371. https://doi.org/10.1016/S0007-1935(17)36887-2

Smith TS, Amstrup SC, Kirschhoffer BJ and York G 2020 Efficacy of aerial forward-looking infrared surveys for detecting polar bear maternal dens. *PLoS One 15*: 1-10. https://doi.org/10.1371/journal.pone.0222744

Špinka M 2019 Animal agency, animal awareness and animal welfare. *Animal Welfare 28*: 11-20. https://doi.org/10.7120/09627286.28.1.011

Stirling I 1974 Midsummer observations on behavior of wild polar bears (*Ursus maritimus*). *Canadian Journal of Zoology 52*: 1191-1198. https://doi.org/10.1139/z74-157

Stirling I 1998 *Polar Bears*. University of Michigan Press: Ann Arbor, MI, USA

Stirling I, Thiemann GW and Richardson E 2008 Quantitative support for a subjective fatness index for immobilized polar bears. *Journal of Wildlife Management 72*: 568-574. https://doi.org/10.2193/2007-123

Swaisgood RR and Shepherdson DJ 2005 Scientific approaches to enrichment and stereotypies in zoo animals: What's been done and where should we go next? *Zoo Biology 24*: 449-518. https://doi.org/10.1002/zoo.20066

Swaisgood RR, White AM, Zhou X, Zhang H, Zhang G, Wei R, Hare VJ, Tepper EM and Lindburg DG 2001 A quantitative assessment of the efficacy of an environmental enrichment programme for giant pandas. *Animal Behaviour 61*: 447-457. https://doi.org/10.1006/anbe.2000.1610

Tetley CL and O'Hara SJ 2012 Ratings of animal personality as a tool for improving the breeding, management and welfare of zoo mammals. *Animal Welfare 21*: 463-476. https://doi.org/10.7120/09627286.21.4.463

Todman JB and Dugard P 2001 *Single-Case and Small-N Experimental Designs: A Practical Guide to Randomization Tests, First Edition*. Routledge: New York, USA. https://doi.org/10.4324/9781410600943

Tuyttens FAM, Stadig L, Heerkens JLT, van Laer E, Buijs S and Ampe B 2016 Opinion of applied ethologists on expectation bias, blinding observers and other debiasing techniques. *Applied Animal Behaviour Science 181*: 27-33. https://doi.org/10.1016/j.applanim.2016.04.019

Vickery S and Mason G 2003a Understanding stereotypies in captive bears: The first step towards treatment. In: Gilbert TC (ed) *Proceedings of The Fifth Annual Symposium on Zoo Research* pp 38-51. Federation of Zoological Gardens of Great Britain and Ireland: Winchester, UK

Vickery SS and Mason GJ 2003b Behavioral persistence in captive bears: Implications for reintroduction. *Ursus 14*: 35-43

Wagman JD, Lukas KE, Dennis PM, Willis MA, Carroscia J, Gindlesperger C and Schook MW 2018 A work-for-food enrichment program increases exploration and decreases stereotypies in four species of bears. *Zoo Biology 37*: 3-15. https://doi.org/10.1002/zoo.21391

Waroff AJ, Fanucchi L, Robbins CT and Nelson OL 2017 Tool use, problem-solving, and the display of stereotypic behaviors in the brown bear (*Ursus arctos*). *Journal of Veterinary Behavior 17*: 62-68. https://doi.org/10.1016/j.jveb.2016.11.003

Watters JV 2014 Searching for behavioral indicators of welfare in zoos: Uncovering anticipatory behavior. *Zoo Biology 33*: 251-256. https://doi.org/10.1002/zoo.21144

Watters JV, Krebs B and Pacheco E 2019 Measuring welfare through behavioral observation and adjusting it with dynamic environments. In: Kaufman A, Bashaw M and Maple T (eds) *Scientific Foundations of Zoos and Aquariums: Their Role in Conservation and Research* pp 212-240. Cambridge University Press: Cambridge, UK. https://doi.org/10.1017/9781108183147.009

Wechsler B 1991 Stereotypies in polar bears. *Zoo Biology 10*: 177-188. https://doi.org/10.1002/zoo.1430100209

Wechsler B 1992 Stereotypies and attentiveness to novel stimuli: A test in polar bears. *Applied Animal Behaviour Science 33*: 381-388. https://doi.org/10.1016/S0168-1591(05)80074-7

Wemelsfelder F and Mullan S 2014 Applying ethological and health indicators to practical animal welfare assessment. *Revue Scientifique et Technique (International Office of Epizootics) 33*: 111-120. https://doi.org/10.20506/rst.33.1.2259

Wenker CJ, Stich H, Müller M and Lussi A 1999 A retrospective study of dental conditions of captive brown bears (*Ursus arctos* spp) compared with free-ranging Alaskan grizzlies (*Ursus arctos horribilis*). *Journal of Zoo and Wildlife Medicine 30*: 208-221

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

Wielebnowski NC, Fletchall N, Carlstead K, Busso JM and Brown JL 2002 Non-invasive assessment of adrenal activity associated with husbandry and behavioral factors in the North American clouded leopard population. *Zoo Biology 21*: 77-98. https://doi.org/10.1002/zoo.10005

Wiig Ø, Amstrup S, Atwood T, Laidre K, Lunn N, Obbard M, Regehr E and Thiemann G 2015 *Ursus maritimus*. The IUCN Red List of Threatened Species: e.T22823A14871490. https://doi.org/10.2305/IUCN.UK.2015-4.RLTS.T22823A14871490.en

Williams E, Chadwick CL, Yon L and Asher L 2018 A review of current indicators of welfare in captive elephants (*Loxodonta africana* and *Elephas maximus*). *Animal Welfare 27*: 235-249. https://doi.org/10.7120/09627286.27.3.235

Wooddell LJ, Hamel AF, Murphy AM, Byers KL, Kaburu SSK, Meyer JS, Suomi SJ and Dettmer AM 2017 Relationships between affiliative social behavior and hair cortisol concentrations in semi-free ranging rhesus monkeys. *Psychoneuroendocrinology 84*: 109-115. https://doi.org/10.1016/j.psyneuen.2017.06.018

Yamanashi Y, Teramoto M, Morimura N, Nogami E and Hirata S 2018 Social relationship and hair cortisol level in captive male chimpanzees (*Pan troglodytes*). *Primates 59*: 145-152. https://doi.org/10.1007/s10329-017-0641-8

Yon L, Williams E, Harvey ND and Asher L 2019 Development of a behavioural welfare assessment tool for routine use with captive elephants. *PLoS One 14*: 1-20. https://doi.org/10.1371/journal.pone.0210783

https://doi.org/10.7120/09627286.30.1.001 Published online by Cambridge University Press