MAMMAL COMMUNICATIONS Volume 10 ISSN 2056-872X (online)

Development of a visual body condition score system for juvenile harbour seals (*Phoca vitulina*) in rehabilitation centres.

Anna McMillan, Michal Zatrak, Richard Ilderton & Robyn Grant

Development of a visual body condition score system for juvenile harbour seals (*Phoca vitulina*) in rehabilitation centres.

Anna McMillan[1], Michal Zatrak[1,2], Richard Ilderton[3] and Robyn Grant[1*]

ABSTRACT

Harbour seals (*Phoca vitulina*) treated at rescue centres have lower survival rates than grey seals (*Halichoerus grypus*). Malnutrition and weight are closely associated with recovery; however, weighing them can be invasive, stressful and poses a risk of injury to staff. This project developed a non-invasive body condition score (BCS) system to monitor the body condition of juvenile harbour seals in rescue centres. Images from CCTV footage and weight data were used to develop the BCS system, which was subsequently tested using online questionnaires from a range of users. Using two different metrics, the developed four-point BCS scale had a fair (Kappa = 0.36) and substantial (Kendall = 0.654, $p < 0.001$) level of inter-rater reliability, and was significantly correlated to body weight (rho = 0.718, s = 164994, $p < 0.001$). This system can be implemented in rescue centres alongside existing practices to improve weight monitoring and reduce the handling of harbour seals, with important implications for seal stress and welfare.

INTRODUCTION

The United Kingdom (UK) is home to ~40% of Europe's harbour seals (*Phoca vitulina*), although several colonies are declining (Thompson *et al.,* 2010). Rehabilitation centres support the conservation of harbour seals, admitting hundreds of individuals annually in the UK (MacRae *et al.,* 2011) with ~57% surviving to release. Most seals admitted are juveniles (<1 year old) with malnutrition. Harbour seal survival odds are 4.6x lower than grey seals (*Halichoerus grypus*) treated at the same facilities (Zatrak *et al.*, 2023). Body weight is key to rehabilitation success (Harding *et al.*, 2005; MacRae *et al.,* 2011; Martinez, 2011; Zatrak *et al.*, 2023); however, juvenile harbour seals struggle to maintain weight in captivity, with many seals in rescue centres being ~200 grams below the usual birth weight (~8 kg) (MacRae *et al.,* 2011).

Rescue centres weigh the animals and take measurements to inform feeding protocols, veterinary treatments and release decisions (Osinga & Hart, 2010). However, these measurements require human contact to restrain the animal (Barnett *et al.*, 2000; MacRae *et al.,* 2011; Martinez, 2011), which is stressful, and affects the animal's welfare and recovery, as well as posing an injury risk to staff (Grogan & Kelly, 2013; Mullineaux, 2014). Therefore, weight measurements are often collected infrequently and there is a need to develop an additional method to monitor weight throughout recovery. When repeated measurements of an animal's condition need to be taken, many facilities will opt for a less invasive visual BCS system (Edmonson *et al.*, 1989; German & Morgan, 2008; Gant *et al.*, 2016; Schiffmann *et al.*, 2017). While some rescue centres adopt a BCS for harbour seals (Martinez, 2011), there is not yet a detailed, published, and validated system for rehabilitators to access. Therefore, this study will develop and validate a new BCS system for juvenile harbour seals.

[1]Department of Natural Science, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, UK

[2]Department of Biological Sciences, University of Chester, Chester, UK

[3]Tynemouth Seal Hospital, Grand Parade, Tynemouth, North Shields, NE30 4JF, UK

*Corresponding author: robyn.grant@mmu.ac.uk

Key words: welfare, pinnipeds, rescue, malnutrition, weight

Full citation: McMillan, A., Zatrak, M., Ilderton, R. & Grant, R. (2024) Development of a visual body condition score system for juvenile harbour seals (*Phoca vitulina*) in rehabilitation centres. Mammal Communications 10: 1-7, Blandford Forum. (DOI: 10.59922/YANS5898)

METHODS

Subjects and Data

Animals used in this study were rehabilitating juvenile harbour seals aged between four and 16 months. Age was determined using the system described in Zatrak *et al.*, (2023). All individuals were housed at Tynemouth Seal Hospital during August 2019 and September 2021. Weights were obtained from rescue centre records. Images of the seals were taken from CCTV footage of the quarantine pens.

Pilot Five Point BCS

Typically, BCS systems use a five point scale, which ranges from severely under- to over-weight (Gant et al., 2016; Mullins et al., 2019; Schiffman et al., 2017). Therefore, we initially designed a five point scale (Supplement A) ranging from <10 kg to >20 kg with 2.5kg weight intervals between each score. Using anatomical markers identified by Warren's (2021) leopard seal study (*Hydrurga leptonyx*), we adopted similar terminology that could be used as text descriptions to guide users in assigning a score. We also included images of seals from CCTV footage as visual aids. This system was tested by twenty-four users (including university students, seal rehabilitation staff, academics researching mammals and members of the public). They completed an online questionnaire designed in Google Forms (Google, 2022), to apply the BCS system to images of ten seals from CCTV footage. Users were also asked to define their experience with BCS and harbour seals (Supplement B).

Four Point Scale

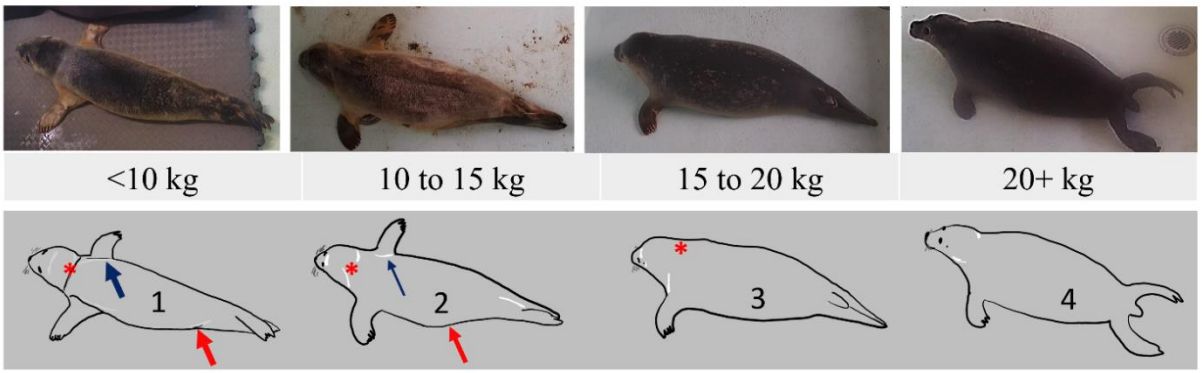
Following feedback and analysis of the five point scale, a revised four point scale was developed. The four point scale used CCTV footage and weight records of seven harbour seals (four female, three males). The final system expanded the weight range (9.4 to 22.5 kg as opposed to 8.3 to 19.9 kg) and divided this into four scores with larger weight intervals (five kg instead of 2.5kg) (Figure 1, and Supplement C). To test the updated scale, an additional questionnaire was designed using Google Forms (Google, 2022). This was distributed to eight users (two academics studying pinnipeds, two university students, and four members of the public), who were asked to score nineteen seals. Care was taken to ensure that the images in this questionnaire had more standard positions that were more comparable to reference images.

Statistical Analysis

To test inter-rater reliability a Fleiss-kappa analysis and Kendall's Test of Concordance, were used and interpreted as follows; 0.01 to 0.20 represented none to slight agreement, 0.21 – 0.40 fair, 0.41 – 0.60 moderate, 0.61 – 0.80 substantial and 0.81 to 1.00 as perfect agreement (recommended by Cohen in McHugh (2012)). A Kruskal Wallis test determined significant differences between scores assigned by users of different levels of experience on the five-point system, but not the four-point due to its smaller sample size.

The accuracy of the BCS systems were tested by comparing the expected BC scores, from the actual weight categories to the scores assigned during the questionnaires. The percentage of answers that matched the expected BC score was calculated, along with over- and under-estimations. The root mean square error (RMSE) was also calculated between the expected and actual scores to provide a measure of error for each system. A Kruskal Wallis test with post-hoc Dunn's test compared differences in weights between scores. A Spearman's Rank test of correlation was used to calculate the strength and direction of the relationship between body mass and body condition scores for both scales.

Figure 1. A summary figure of the Four Point BCS system developed for juvenile harbour seals, with example seal photographs. As well as the general silhouette of the animal, skull shape, the prominence of the neck (red asterisk), shoulder blades (blue arrow) and hip bones (red arrow) are all used to judge body condition.



RESULTS

Inter-rater reliability of scores

The five point and four point scales had "fair" scores of agreement of 0.217 (K = 0.217, Z = 22.3, p < 0.001) and 0.364 (K = 0.364, Z = 14.2, p < 0.001), respectively, with the four-point BCS having a slightly higher K-score. The Kendall test of concordance reported "substantial" levels of agreement in both the five point scale (W = 0.725, chisq = 157, p < 0.001) and the four point scale (W = 0.654, chisq = 94.2, p < 0.001).

There was no significant difference in user scores of the five point BCS with different levels of experience with harbour seals (Kruskal Wallis: chi = 0.809, df = 3, p = 0.847) or BCS systems (chi = 2.041, df = 3, p = 0.564).

Accuracy of the BCS Scale

22.5% of scores were a direct match to the weight categories in the five point BCS scale, and 49.67% of scores in the four point BCS Scale. The four point scale had a lower error (RMSE = 0.86) than the five point scale (RMSE = 1.47) (Table 1). Accuracy scores were omitted for score three on the five-point scale since none of the seals weights in the pilot study were in the third category.

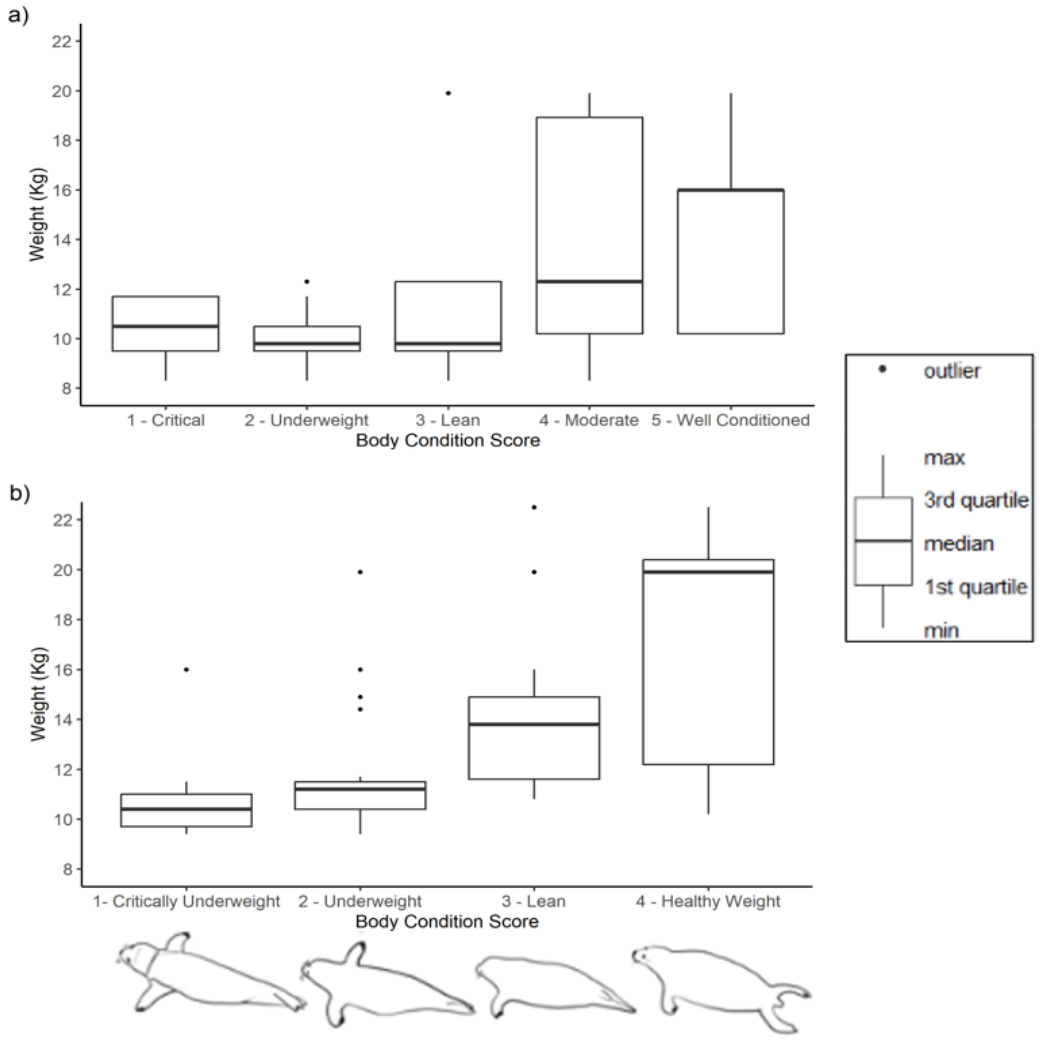
Table 1. Table of BCS accuracy for the Five and Four Point Scales and Root Mean Square Error (RMSE) for the scores. N Ass = not assessed. NA = not applicable.

Method		Five Point Scale	Four Point Scale
Overall Scoring Accuracy	*Correct*	54/240 (22.5%)	76/152 (49.67%)
	Over	145/240 (60.42%)	57/152 (37.5%)
	Under	41/240 (17.08%)	19/152 (19%)
Score RMSE		1.47	0.856
Score One	*Correct*	18/96 (18.75%)	12/16 (75%)
	Over	78/96 (81.25%)	4/16 (25%)
	Under	0 (0%)	0/16 (0%)
Score Two	*Correct*	23/96 (23.96%)	43/104 (41.35%)
	Over	51/96 (53.13%)	48/104 (46.15%)
	Under	22/96 (22.92%)	13/104 (12.5%)
Score Three	*Correct*		7/16 (43.75%)
	Over	N Ass	5/16 (31.25%)
	Under		4/16 (25%)
Score Four	*Correct*	8/24 (33.33%)	14/16 (87.5%)
	Over	16/24 (66.66%)	0/16 (0%)
	Under	0/24 (0%)	2/16 (12.5%)
Score Five	*Correct*	5/24 (20.83%)	
	Over	0/24 (0%)	NA
	Under	19/24 (79.16%)	

There were significant differences in weight (kg) between the categories of the five point BCS Scale (Kruskal-Wallis: chi-squared = 53.064, df = 4, p-value < 0.001), and four point BSC scale (chi-squared = 78.338, df = 3, p-value < 0.001). However, Figure 2 showed much more overlap of scores in the five point scale (Figure 2a), compared to the four point scale (Figure 2b), indicating that consecutive scores were hard to separate. Post Hoc tests illustrated that, for the five point scale, there were significant differences (p < 0.01) between many scores, apart from one and two, one and three, two and three, and four and five. In comparison, all scores were significantly different on the four point scale (p < 0.01), apart from one and two and three and four.

A Spearman's Rank Correlation found a significant positive correlation between body mass (kg) and body condition scores for both the five point BCS Scale (rho = 0.405, df = 238 , p < 0.001) and the four point BCS Scale rho = 0.718, df = 150, p < 0.001), with the four point scale having a higher coefficient (rho).

Figure 2. Box Plots depicting the weights (Kilograms) of harbour seals assigned to each body condition score for Five Point (a) and Four Point Scales (b). Example body shapes can be seen below the Four Point Scale



DISCUSSION

Our study developed a novel four point BCS scale for juvenile harbour seals that is now validated, with a good correlation between assigned scores and body weight and a fair level of agreement between raters. Intended to be used alongside existing methods, this new system will provide rescue centres with an additional method of monitoring seal recovery, allowing them to prioritise animal welfare by minimising weighing frequency without sacrificing the ability to monitor weight, which is an important factor related to survival and release (Harding *et al.*, 2005; MacRae *et al.,* 2011; Martinez, 2011; Zatrak *et al.*, 2023).

Following feedback from the users of the five point scale, the new four point scale was adapted to have fewer scores with larger weight intervals because users struggled to discern between categories when weight was close to a boundary. Whilst other BCS systems have implemented half points on the scale to account for this (Fernando *et al.*, 2009), our solution maintained the simplicity and practicality of the system, in agreement with Schiffmann *et al.*, (2017).

The four point scale gave an improvement in reliability (from kappa = 0.217 to 0.364), with reduced error (RMSE = 1.47 to 0.86). A kappa score of larger than 0.35 denotes a level of agreement well above that of random chance (McHugh, 2012); however, it is below the levels often seen in other BCS systems. For instance, Gant *et al.* (2016) reported a kappa score of 0.41 for their domestic dog BCS method, which increased to 0.70 in experienced raters. Further user training and experience with the system may increase our reliability score. However, prior experience of users did not significantly affect the reliability of the five point scale, which supports the feasibility of the application of this system in a rescue centre setting, where user experience can be varied (Grogan *et al.,* 2013; Mullineaux, 2014). Training of our BCS scale may improve its reliability in practice (Vieira *et al.*, 2015), and we recommend that should happen before

adoption in rescue centres. Specifically, the system should be tested on seals in-person in the same manner the system would be applied in practice. Indeed, sample size was a limiting factor of the study, and we would advise further validation of the system with a larger and more diverse sample size to better account for individual variability, due to factors such as age, sex or life history events.

The challenging body plan of pinnipeds, with their thick layers of blubber, can mask body condition markers (Schiffmann et al., 2017) and may reduce the reliability of the scoring system. However, there are existing validated BCS systems for species with similar body shapes, such as common dolphins (*Delphinus delphis*) (Joblon et al., 2014), magellanic penguins (*Spheniscus magellanicus*) (Clements & Sanchez, 2015) and leopard seals (*Hydrurga leptonyx*) (Warren, 2021); therefore, body shape should not limit the application of the BCS system to harbour seals.

Our BCS system could be expanded to include older seals, and for other uses, such as in field settings (e.g during aerial surveys assessing colony health (Warren, 2021; Wall et al., 2023)) or zoos (Clegg & Butterworth, 2017). Although further testing and validation would be needed for these additional use-cases. For application to juvenile harbour seals in rescue centres, we are confident in the reliability and accuracy of our developed four point BCS system, which we have made available in Supplement C.

ACKNOWLEDGEMENTS

We would like to extend our gratitude to Tynemouth Seal Hospital for allowing us to use their records and footage which formed the basis of this project, as well as for their assistance in testing and validating the system. We would like to also thank the thirty-two participants who reviewed the body condition score systems. Thanks also go to the two Reviewers who gave us feedback on our manuscript.

REFERENCES

Barnett, J.E.F., Woodley, A.J., Hill, T.J. & Turner, L. (2000) Conditions in grey seal pups (*Halichoerus grypus*) presented for rehabilitation. *Veterinary Record,* 147: 98-104.

Clegg, I. L., & Butterworth, A. (2017) Assessing the welfare of pinnipeds. In: Butterworth, A. (Ed.) Marine mammal welfare: Human induced change in the marine environment and its impacts on marine mammal welfare, 273-295. Cham, Springer International Publishing.

Clements, J. and Sanchez, J.N., 2015. Creation and validation of a novel body condition scoring method for the Magellanic penguin (*Spheniscus magellanicus*) in the zoo setting. *Zoo Biology* 34(6): 538-546.

Edmonson, A.J., Lean, I.J., Weaver, L.D., Farver, T. & Webster, G. (1989) A body condition scoring chart for Holstein dairy cows. *Journal of Dairy Science* 72: 68-78.

Fernando, P., Janaka, H.K., Ekanayaka, S.K., Nishantha, H.G. & Pastorini, J. (2009) A simple method for assessing elephant body condition. *Gajah* 31: 29-31.

Gant, P., Holden, S.L., Biourge, V. & German, A.J. (2016) Can you estimate body composition in dogs from photographs? *BMC Veterinary Research* 12: 1-12.

German, A.J. & Morgan, L.E. (2008) How often do veterinarians assess the bodyweight and body condition of dogs? *Veterinary Record* 163: 503-505.

Grogan, A. & Kelly, A. (2013) A review of RSPCA research into wildlife rehabilitation. *Veterinary Record* 172: 211-211.

Harding, K.C., Fujiwara, M., Axberg, Y. & Härkönen, T. (2005) Mass-dependent energetics and survival in harbour seal pups. *Functional Ecology* 19: 129-135.

Joblon, M.J., Pokras, M.A., Morse, B., Harry, C.T., Rose, K.S., Sharp, S.M., Niemeyer, M.E., Patchett, K.M., Sharp, W.B. and Moore, M.J., 2014. Body condition scoring system for delphinids based on short-beaked common dolphins (*Delphinus delphis*). *J. Mar. Anim. Ecol* 7(2): 5-13.

MacRae, A.M., Haulena, M. & Fraser, D. (2011) The effect of diet and feeding level on survival and weight gain of hand-raised harbor seal pups (*Phoca vitulina*). *Zoo biology* 30: 532-541.

Martinez, C.L., 2011. Analysis of Pacific Harbor Seal Survival at Wolf Hollow Wildlife Rehabilitation Center, Friday Harbor, WA. Available at: https://digitalcommons.calpoly.edu/ascisp/6/ (Accessed: 12 October 2023).

McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* 22: 276-282.

Mullins, I. L., Truman, C. M., Campler, M. R., Bewley, J. M., & Costa, J. H. (2019) Validation of a commercial automated body condition scoring system on a commercial dairy farm. *Animals* 9: 287.

Mullineaux, E. (2014) Veterinary treatment and rehabilitation of indigenous wildlife. *Journal of Small Animal Practice* 55: 293-300.

Osinga, N., & Hart, P. (2010) Harbour seals (Phoca vitulina) and rehabilitation. *NAMMCO Scientific Publications* 8: 355-372.

Schiffmann, C., Clauss, M., Hoby, S. & Hatt, J.M. (2017) Visual body condition scoring in zoo animals–composite, algorithm and overview approaches. *Journal of Zoo and Aquarium Research* 5: 1-10.

Thompson, D., Duck, C. & Lonergan, M.E. (2010) The status of harbour seals (*Phoca vitulina*) in the United Kingdom. *NAMMCO Scientific Publications* 8: 117-127.

Vieira, A., Brandão, S., Monteiro, A., Ajuda, I. & Stilwell, G. (2015) Development and validation of a visual body condition scoring system for dairy goats with picture-based training. *Journal of Dairy Science* 98: 6597-6608.

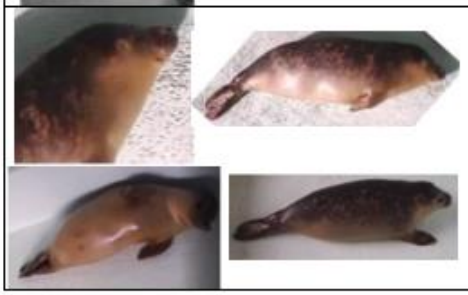
Warren, J. (2021) A comparison of statistical techniques for evaluating body condition in New Zealand leopard seals (Hydrurga leptonyx) using citizen science data (Doctoral dissertation, University of York). Available at: https://etheses.whiterose.ac.uk/28944/ (Accessed: 12 October 2023).

Wall, D., Thalmann, S., Wotherspoon, S., & Lea, M. A. (2023). Is regional variability in environmental conditions driving differences in the early body condition of endemic Australian fur seal pups? *Wildlife Research* 50: 993-1007.

Zatrak, M., Brittain, S., Himmelreich, L., Lovick-Earle, S., Pizzi, R., Shaw, K.J., Grant, R.A. & Geary, M. (2023) Factors affecting the survival of harbor (*Phoca vitulina*) and gray seal (*Halichoerus grypus*) juveniles admitted for rehabilitation in the UK and Ireland. *Marine Mammal Science* 392: 462-480.

SUPPLEMENTARY INFORMATION FOR DEVELOPMENT OF A VISUAL BODY CONDITION SCORE SYSTEM FOR JUVENILE HARBOUR SEALS (*PHOCA VITULINA*) IN REHABILITATION CENTRES.

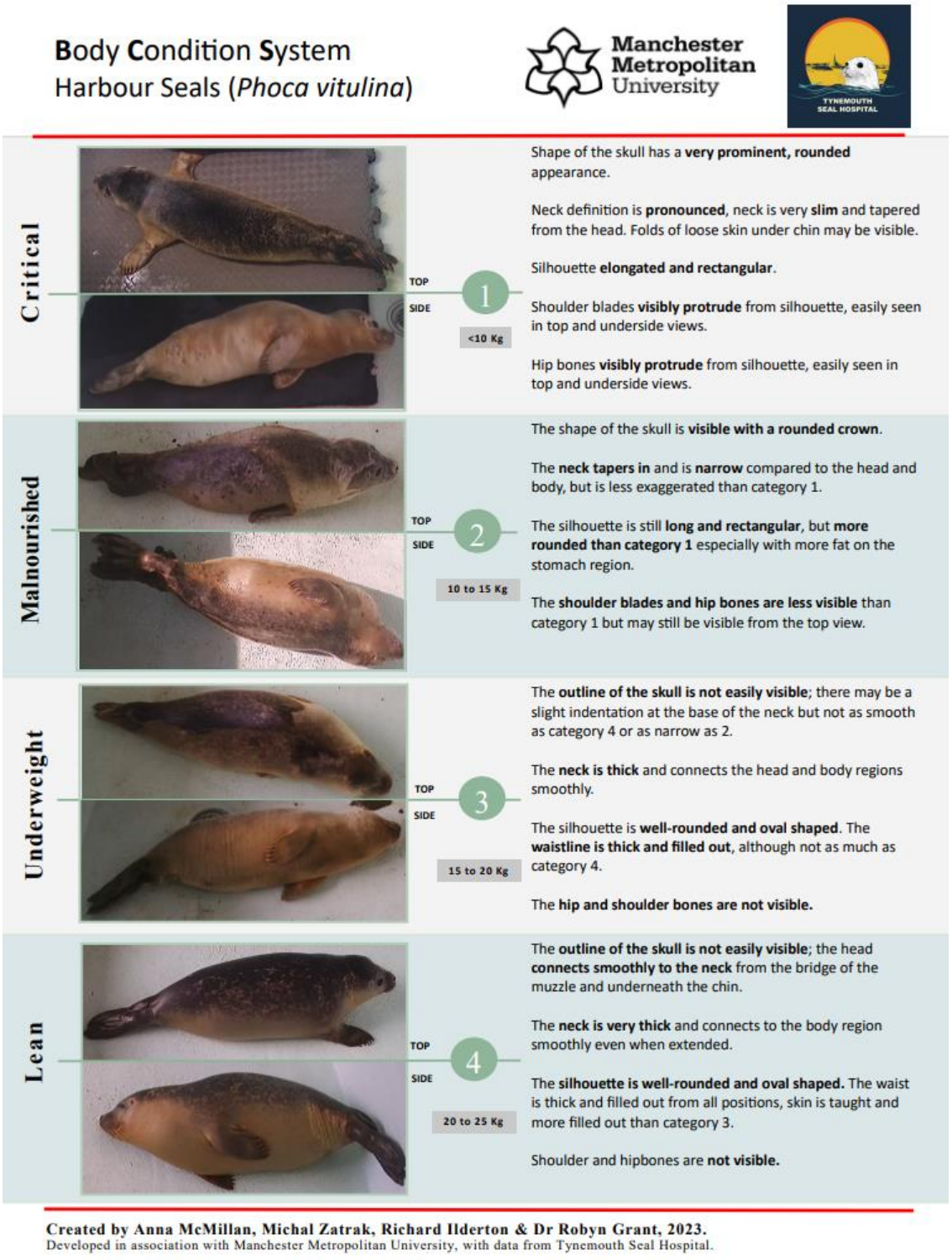
Supplement A: The Five Point BCS System

	**Score One ~ Severely Underweight** In the head view the shape of the **skull** is easily visible, with a rounded appearance most evident at the crown of the head. The **neck** is very narrow and tapered from the skull, when extended this is most evident. Folds of loose skin are visible underneath the chin. The overall **silhouette** is elongated and rectangular. **Shoulder blades and hipbones** visibly protrude from the silhouette and are easily identifiable in top and underside view.
	**Score One ~ Underweight** As seen in score one, the shape of the **skull** is visible, with a rounded appearance. The **neck** is narrow and defined, but to a lesser degree than score one. There is more fat deposited at the base of the skull and underneath the chin. There are still folds of loose skin under the chin. The **silhouette** is still elongated and rectangular, but there is slightly more fat on the stomach region than score one. The **shoulder blades and hipbones** visibly protrude from the silhouette but are not as sharp as in score one.
	**Score three ~ Lean** The **skull** is not as rounded as score two, starting to take on a smoother continuum with the neck area. The **neck** is thicker than score two, merging the head and body regions smoothly with a small indentation at the base of the skull. The **silhouette** is taking on a more rounded appearance, with fat deposition on the stomach and flanks filling out the body. The stomach region appears taunt and full. The **shoulder bones and hips** can still be observed but are much more subtle than in score two.
	**Score Four ~ Moderate** The outline of the **skull** is not easily visible, the head region connects the bridge of the muzzle to the neck in a smooth continuum. The **neck** is thick and the head and body regions connect smoothly. The neck does not dip inwards. The **silhouette** is well-rounded and oval shaped. The waistline is thick and filled out in all positions. Stomach is taunt. The **shoulder bones and hipbones** are not visible.
	**Score Five ~ Well Conditioned** The outline of the **skull** is not easily visible; the head connects smoothly to the neck from the bridge of the muzzle and underneath the chin. The **neck** is very thick and connects to the body region smoothly, even when extended. The **silhouette** is well rounded and oval shaped. The waistline is thick and filled out from all positions. Stomach is taut and full. Compared to score four, the seal is more filled out in all body regions. **Shoulder and hipbones** are not visible.

Supplement B: Questionnaire Responses pertaining to the experience and backgrounds of the users testing the Five Point Scale (Pilot Study).

User Seal Experience	% of Users	User BCS system experience	% of Users
No Experience with harbour seals	62.5% (15/24)	Never heard of BCS systems	16.7% (4/24)
Basic Experience – a few weeks to six months working with/on this species.	16.7% (4/24)	Have heard of BCS but have no practical experience	29.2% (7/24)
Above average – 6 to 12 months experience	4.2% (1/24)	Aware of BCS but limited practical experience.	37.5% (9/24)
Expert/Advanced – equivalent to over a year of experience working with/on this species.	16.7% (4/24)	Deeply familiar and experienced with BCS systems.	16.7% (4/24)

Supplement C: The Four Point BCS System



Created by Anna McMillan, Michal Zatrak, Richard Ilderton & Dr Robyn Grant, 2023.
Developed in association with Manchester Metropolitan University, with data from Tynemouth Seal Hospital.