An assessment of the welfare of non-human primates used in neuroscience research

Rhyanne Heppenstall

Department of Experimental Psychology

Linacre College

University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2023

Declaration

This thesis has only been submitted to the University of Oxford for the degree of Doctor of Philosophy. No part has been accepted or is currently being submitted for any other degree or other qualification in this University or elsewhere.

Funding for this project was provided by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship and a Wellcome Trust Strategic Award. This work was conducted via the Interdisciplinary Bioscience Doctoral Training Centre in the form of an Industrial Cooperative Awards in Science and Technology (iCASE) partnership, in collaboration with Vetronic Services Ltd (now part of Burtons Medical Equipment Ltd).

Word count: 58,037

Contributions

I declare that this thesis is all my own (RH) work except when explicitly stated below.

Chapter 3 -The impact of general anaesthesia on non-human primates

The concept of the study was conceived by RH and Caroline Bergmann (CB). This chapter was supervised by CB, Mark Walton (MW) and Michael Bonsall (MB). RH conducted the majority of the data collection but was assisted by Katie Collier (KC) and Stuart Mason (SM) (research assistants). Research assistant Rhiannon Austin (RA) conducted all analyses and produced graphs and figures, with input from MB. RA and RH wrote methods and results section. RH, RA, and CB interpreted results. RH wrote all other components of the chapter.

Chapter 4 -Automated behavioural assessment using accelerometers

The study concept was conceived by CB, RH, and DPhil student Natasha Gillies (NG). This chapter was supervised by CB, MW, and Maarten de Vos (MdV). A pilot study to create an initial model and obtain proof of concept was conducted by NG. Video and actigraphy data were collected by RH and NG. Behavioural observations were conducted by RH, NG, and KC. RH continued model development and validation. RH interpreted results. RH wrote all components of the chapter and created all tables and figures.

Chapter 5 -Cortisol and non-human primate welfare

The study concept was conceived by CB and RH. Faecal sample collection was a collaborative effort between RH and the animal technicians, with assistance from KC and SM. Hair samples were primarily collected by RH in collaboration with neuroscience researchers but assisted by KC and SM. Faecal samples were assayed for cortisol at Deutsches Primatenzentrum Endocrinology lab, Gottingen, Germany. Hair samples were assayed for cortisol at the Faculty of Psychology, TU Dresden, Germany. Shipment and storage logistics were conducted by RH and CB. Faecal data analysis, graphs, and interpretation of results was by RH. Hair cortisol analysis was conducted by RH in collaboration with Dr Esther Carlitz (EC), TU Dresden. Graphs and interpretation of results was by RH. RH wrote all components of the chapter under supervision from CB, MW, and EC.

Chapter 6 -Development of a novel ECG device for measuring heart rate variability

The study concept was conceived by CB and Dr Keith Simpson (KS), Vetronic Services Ltd. Technical development of the ECG was by KS. ECG trials were conducted by RH and CB. RH interpreted results. RH wrote all components of the chapter and created all figures, under the supervision of CB, MW, and KS.

Statement about the COVID–19 pandemic

The work undertaken for this thesis was subject to disruption caused by the COVID-19 pandemic in 2020. Imposed lockdowns abruptly stopped ongoing, longitudinal data collection protocols in March 2020. Due to the vulnerability of non-human primates (NHPs) to COVID-19, access to the animal facility was restricted to essential personnel only (e.g. animal care staff) and all research with the NHPs (both neuroscience and welfare) was initially prevented to minimise risk to the animals. The halt to neuroscience testing ultimately impacted on this work, as welfare data were collected alongside neuroscience protocols. The nature of this work required continuous, longitudinal assessment and so the cessation of all work disrupted any benefit of further data collection. At the time of the pandemic, the ECG project was at a stage where trials on animals under general anaesthesia was required. The halt to neuroscience protocols delayed any planned general anaesthetics by several months, both due to restricting contact with NHPs but also because such procedures required personnel to be in closer contact. I was granted a 12-week extension to work on this project, however some work has had to be adjusted due to the loss of data.

Ethics statement

Non-human primates discussed in this thesis were enrolled in neuroscience studies under project licences (PPL) issued by the Secretary of State in accordance with the Animals (Scientific Procedures) Act 1986 (Amendment Regulations 2012) - A(SP)A. All PPLs underwent ethical review by the local Animal Welfare Ethical Review Board (AWERB) and Animals in Science Committee (ASC). All methods conducted as part of this thesis were carried out by appropriately trained persons. To minimise welfare impact, all methods were considered to be below the 'lower threshold' defined in A(SP)A and were not considered to be regulated procedures. However, in order to collect some welfare parameters, cooperation with regulated procedures was required. Where a regulated procedure was applied to a protected animal, such as restraint in a primate chair, in accordance with the relevant PPL, this was conducted by a suitably trained person holding an appropriate personal licence (PIL).

Acknowledgements

This thesis would not have been possible without the enthusiasm, encouragement, and guidance of my supervisor, Caroline Bergmann. Your continued passion for animal welfare inspires me to keep going and always re-ignites my love for the topic. Thank you!

A huge thank you to all those who have worked in a supervisory role towards my chapters and the thesis as a whole, your guidance has been invaluable; John Duncan, Maarten de Vos, Michael Bonsall, Esther Carlitz and Keith Simpson. Particular thanks to Mark Walton for helping to keep me level-headed at the end with his swift responses and supportive feedback.

To those that passed through our welfare lab in various roles. Thank you for being assistant collar stitchers, barbers, poo organisers and every other strange role this work has led us into – Stuart Mason, Anna Mitchell, Katie Collier, Angus Fisk, Charlie Johnson and Jess Lock.

For their invaluable assistance with data analysis, statistical advice, pilot studies and general help with navigating PhD life and it's challenges; Rhiannon Austin, Natasha Gillies, Thomas Rawson and Sofia Minano Gonzalez. Thank you.

I can never say enough thank yous to all of the macaques involved in this study! Also, to the neuroscience researchers that have never seemed to mind my persistent requests and constant hovering. Without the hard work and dedication of all the animal care staff and vets, this thesis would not have been possible. Your commitment to the welfare of your animals is honourable and thank you for always doing your best to accommodate my work. Particularly to Andrew Emberton, Maria Martinez, Katie Underdown and Kelly Simpson who have had to field through many of my emails and let me pick their brains about records and animal training regimes.

A thank you to the staff at Hilltop veterinary practice and the sheep researchers at the University of Nottingham, who didn't mind me constantly getting in their way with my ECG!

I could not have got by without my friends, who have kept on checking in and making sure I am staying sane throughout this whole process: Dan and Emma, Sarah and Dean, Drew and Rich. Particularly, Lindsey Taylor and Alice May, who have listened to me moan and helped me navigate the balance of work and new parenthood.

To all of my family, a huge thank you! To my Mum and Dad, your support has got me to where I am and you will never know how grateful I am to have you. Special thank you for all of the cups of tea provided as fuel at the end. I promise to stop going on about monkey welfare and venting about being a student from now on. Thank you to the rest of my family, who are so fed up with me by now that they 'feel like this is basically our PhD now too,' as my cousin Kim so affectionally put it! A further thank you to my sister-in-law Cheryl for her rapid but thorough proof-reading skills towards the end.

For his dedication to science, thank you to my German shepherd, Arnie. He worked as a perfect model for my ECG and managed to sleep through most of his trials. Now warmly coined 'science dog' in our house for his incredible hard work.

Thank you to my son, Alvin. Despite your arrival making this PhD infinitely more challenging, it has also made it infinitely more rewarding. Thank you for being my joy in the toughest of movements and for making me smile every day. Although all you know about monkeys at the moment is that they go 'oo oo oo,' I promise you there is a lot more (that you will unfortunately be obliged) to learn. Also, thank you for the following contribution provided shortly after a nap (during which I usually tried to get as much work done as possible): 'CbhhhhhAhg0068hmmmmmmmmmmmmmmmmmcdcxb!X43333`836333333333333333IUUUUUUUUUUUUUUUUUUUUG.'

Last but certainly not least, the biggest of thank yous has to go to my husband, Stuart. You have supported me emotionally, financially, and in every single way throughout this entire PhD journey. You have taken on the role of primary cook, cleaner, shopper, parent, and everything else to get me through this. I'm so glad to have you by my side always and I cannot wait to have more time to spend as a family now that this stage is complete. I love you always and I dedicate this thesis to you.

Table of contents

Abstract

The Animals (Scientific Procedures) Act 1986 governs the use of animals in scientific research in the UK. Embedded within this is a requirement to implement the 3Rs: replacement, reduction, and refinement; a key mechanism for minimising the pain, suffering, distress, and lasting harm of research models. To adhere fully to these principles, it is imperative to assess animal welfare. Non-human primates (NHP) human similarities make them both an essential biomedical research model and a species particularly vulnerable to welfare challenges. This thesis investigates non-invasive, objective methods of welfare assessment and applies them alongside neuroscience research to monitor the welfare of rhesus macaques (*Macaca mulatta*).

Accelerometers monitored changes in activity levels following general anaesthesia (GA) and revealed a decline in activity for 5.7 days following GA alone. Additionally, 2.6 days of activity decline were observed after surgery under GA, followed by an activity increase, suggesting post-operative behavioural change. Additionally, accelerometers were used to create a rule-based model for automated behavioural assessment. This model groups macaque activity into 5 species-typical behavioural categories with an overall 69% accuracy.

Physiological welfare parameters were assessed by detecting cortisol in faeces and hair. Faecal cortisol levels were significantly elevated for several days following GA, with a longer and more profound increase after surgery. Hair samples are a valuable measure of chronic stress and cumulative experience, facilitating longitudinal cortisol assessment without influence from transient stressors. Hair cortisol levels were significantly elevated post-surgery and, in some cases, following social disruption.

A customised, neck-based ECG was designed to monitor heart rate variability. The ECG was trialled on anaesthetised and restrained NHPs, dogs, and sheep. Proof of concept was achieved, with R-waves detectable in all species.

Overall, this thesis demonstrates how welfare and neuroscience research can be conducted in parallel for a better understanding of the life-time experience of animal models. This is an essential step towards implementing refinements and improving welfare, which may ultimately improve public perception and help to form the evidence-base required to inform and drive policy change.

Abbreviations

3Rs – Replacement, reduction, and refinement

ACTH – Adrenocorticotrophic hormone

ANS – Autonomic nervous system

AR – Active rest behaviour

ARRIVE guidelines – Animal Research: Reporting of *In Vivo* Experiments

ASC – Animals in Science Committee

AWA - Animal Welfare Act 2006

A(SP)A – Animals (Scientific Procedures) Act 1986

AWERB – Animal Welfare and Ethical Review Body

BVAAWF – British Veterinary Association Animal Welfare Foundation

CRH – Corticotropin-releasing hormone

ECG – Electrocardiogram

EMG – Electromyography

EU SCHEER – European Union Scientific Committee on Health Environmental and Emerging Risks

FCM – Faecal cortisol metabolite

FGM – Faecal glucocorticoid metabolite

FRAME – Fund for the Replacement of Animals in Medical Experiments

GA – General anaesthetic

GC – Glucocorticoid

GL – General locomotion behaviour

HCC – Hair cortisol concentration

HF – High frequency

HPA – Hypothalamic-Pituitary-Adrenal

HR – Heart rate

HRV – Heart rate variability

IA – Inactive behaviour

IBI – Inter-beat interval

LF – Low frequency

MRI – Magnetic resonance imaging

NACWO – Named animal care and welfare officer

NC3Rs – National Centre for the 3Rs

NHP – Non-human primate

NN interval – distance between consecutive, normal R waves in measures of heart rate variability.

NRT – Negative reinforcement training

NVS – Named veterinary surgeon

P – Pacing behaviour

PIL – Personal licence

PNS – Parasympathetic nervous system (sometimes referred to as vagal)

PPG - Photoplethysmography

PPL – Project licence

PQRST - Refers to all waves in an ECG trace (P wave, QRS complex and R wave)

PRT – Positive reinforcement training

PRV – Pulse rate variability

PSDLH – Pain, suffering, distress or lasting harm

QRS – Refers to the Q wave, R wave and S wave of an ECG trace

R – Rapid movement behaviour

RR interval – distance between consecutive R waves in measures of heart rate variability.

RSPCA – Royal Society for the Prevention of Cruelty to Animals

SCHEER – Scientific Committee on Health Environmental and Emerging Risks

SD – Standard deviation

SN – Sinoatrial node

SNS – Sympathetic nervous system

UAR – Understanding Animal Research

UFAW – Universities Federation for Animal Welfare

UGM – Urinary glucocorticoid metabolite

Chapter 1 - Introduction

1.1 Animal welfare and the need for a scientific approach

Considering the welfare of animals is essential in a world where humans are using them for our own gain; in the form of food, entertainment, or as scientific research models. As humans, our behaviours are influencing the lives of many animal species and it is our ethical responsibility to minimise their suffering and promote a good quality of life. Optimising welfare standards can offer benefits to people and the wider environment too. Healthier animals are less likely to transmit disease and will provide higher quantity and quality of produce for us to eat (Dawkins, 2012). Furthermore, the quality of scientific experiments is improved by good welfare in animal research models (Prescott & Lidster, 2017). Preventing harm and promoting animal welfare is also a legal requirement for any persons in the UK 'responsible for' animals as dictated by the AWA (Animal Welfare Act, 2006); which includes all vertebrates, such as pets and domestic species or the A(SP)A (Animals (Scientific Procedures) Act, 1986); including vertebrate and cephalopod species used in scientific experiments.

Prior to the 1800s, philosophers generally considered animals to have only instrumental value, being so distinct from human beings that they could be used by humans in any way we desired. For example, René Descartes (1596-1650) famously referred to animals as 'automata' or machines, believing a lack of language and reasoning suggested that animals were incapable of conscious feelings (Duncan, 2019). Jeremy Bentham in 1823 rejected these and similar claims by expressing a view that animals do have intrinsic value and famously stated 'The question is not, Can they reason? Nor, Can they talk? But, Can they suffer?' (Bentham, 1823).

The importance of animal welfare is rapidly gaining popularity within the general public, as human perceptions of animals develop and pull at our moral compasses. The publication of 'Animal Machines' in the 1960s (Harrison, 1964), which alerted the public to the true experiences

of intensively farmed animals, began this shift in perception and a gradual increase in concern for animal welfare has been seen since. This change in attitudes is reflected in the implementation and advancement of the laws that govern animal use; in farms, as pets, and as biomedical research models. From the UK's first animal protection law in 1822, the Act to Prevent the Cruel and Improper Treatment of Cattle, to the currently employed AWA (Animal Welfare Act, 2006). Furthermore, the publication of the Brambell report (Brambell, 1965) by the British Government and its influence in driving the 'Five Freedoms' concept (see Section 1.1.2) lead to the launch of animal welfare as a scientific discipline. The Brambell report emphasises a need for a scientific approach to welfare assessment, through exploration of various disciplines, such as animal physiology and behaviour. The use of scientific research to assess animal welfare is essential to provide a true understanding of animal well-being. Although it can be valuable to empathise with animals and is usually what drives us to care about those animals in the first place, it cannot be assumed that animals share the same feelings and needs as humans do. We have different evolutionary histories and purposes in life, meaning we may not experience the world in the same way (Hemsworth et al., 2015; Hubrecht, 2014). To be able to correctly assess welfare of non-human animals, scientific research is required to establish what an animal wants and needs.

1.1.1 Defining animal welfare

Although a topic of growing interest, the scientific assessment of the welfare of animals is extremely complex. The first hurdle lies in being able to get researchers to agree on a definition about what animal welfare actually is. Definitions of 'animal welfare' such as that by Professor Donald Broom (Broom, 1988) propose that the 'welfare of an individual is its state as regards its attempts to cope with its environment.' While this is a logical and well accepted definition there are several components that may contribute to the animal's state. This is where conflicting views can often arise, causing disagreement among scientists who view the importance and contribution of these factors to welfare differently.

There are three divergent views of animal welfare currently recognised with focus on biological function, affective state, and natural state (Fraser, 2003, 2008). Assessing an animal's general health can provide a definitive measure of 'biological function' and establishes a clear framework for improvement. For example, by monitoring chicken health in different caging improvements to cage design could be highlighted that would minimise the occurrence of health issues, such as foot and neck lesions (Tauson, 1998). However, many welfare researchers would agree that poor welfare is not always reflected in an animal's physical condition. The 'affective state' view considers a reduction in animal welfare may arise instead from a compromised mental state and an individual's feelings, such as anxiety and stress (Mason & Mendl, 1993). The 'natural state' view implies that animals must be allowed to live naturally and be able to engage in their full behavioural range in order to experience optimum welfare. Most commonly, welfare is considered to be a combination of all of these aspects. However, these three views can often conflict and create difficulties when deciding which measure would be most useful to gain an objective and quantitative welfare assessment (Fraser, 2008). As there is no single 'gold-standard' measure of welfare, it is common, and thought to be good practice among animal welfare scientists to use a combination of methods to explore multiple welfare parameters including measures of animal health, behaviour, and physiology (Broom, 1988; Dawkins, 2006).

1.1.2 Animal sentience and affective state

As early as the 1800s, Charles Darwin proposed that many species share similar emotional experiences to humans, such as fear, pain, and pleasure (Darwin, 1873). He even considered that the expression of these emotions could be similar, for example by differing facial expression and vocalisations. Feelings and emotions, both positive and negative, are often referred to as affective states. The contribution of an animal's affective state to their welfare makes assumptions that animals are conscious beings and do have the capacity to feel. Animal sentience refers to the ability of animals to be able to express feelings and emotions, which can be both positive, such as

pleasure, and negative, e.g. fear or anxiety (Duncan, 2006; Proctor et al., 2013). Absolute proof of sentience in animals is extremely difficult to obtain as this would require a quantitative measure of subjective states. In humans, we rely on the reporting of personal feelings as scientific evidence for human sentience. However, this may still provide unreliable evidence, for example through false reports or an inability to articulate an internal experience. Therefore, understanding the emotions of animals poses even greater challenge, to whom we cannot directly communicate and ask about their subjective experiences. However, growing evidence has implied sentience in many animal species. Although how animal emotions map to human ones is still unclear (de Waal, 2011), it is generally now accepted that all vertebrate species can experience emotions and feelings. This is reflected in the introduction of the Animal Welfare (Sentience) Bill by the UK government, which formally recognises vertebrate animals as sentient beings in domestic law (Animal Welfare (Sentience) Act, 2021). In addition, there is mounting evidence to suggest that many invertebrate species, such as cephalopods and crustaceans, also have at least the ability to sense pain (Proctor, 2012).

The degree to which an animal may be sentient is often contributed to by several factors. It is a popular assumption that animals with more similarities to humans will have a higher degree of sentience (Broom, 2014). This is implied in physiological similarities, such as the presence of comparable nervous systems and neural networks that detect pain in humans, and in cognitive capabilities, such as the ability to form memories. For example, the presence of episodic memory could allow animals to recall aversive events and therefore suffer for longer periods (Bateson, 2004; Mendl & Paul, 2004). However, emotional states such as suffering are subjective experiences, so we can never be certain that other individuals, let alone other species, have the exact same experiences that we do. Even despite the evidence, it is generally the view of welfare scientists that sentience should be assumed, as this provides the safest option to avoiding unnecessary suffering (Broom, 2014).

Assuming sentience of animal species does emphasise a vital need to consider affective state alongside physical well-being when evaluating welfare. The evaluation of psychological health is often required to allow a conclusive estimate of animal welfare, as reflected in changes to welfare frameworks and laws. For example, the 'five freedoms' concept that helped shape the field of animal welfare science has been reformulated in response to updated views of what constitutes good welfare. The freedoms were initially formulated for considering farm animal welfare but are still considered when outlining policies associated with animal welfare and have influenced current welfare legislation (e.g. the Animal Welfare Act 2006) (Mellor, 2016). The five freedoms outline aspirations for animal care that are under human control and can be managed in order to optimise welfare. These were the freedom from hunger or thirst, freedom from discomfort, freedom from pain, injury, or disease, freedom to express normal behaviour, and freedom from fear and distress. Using the freedoms as a basis, a new model was proposed in 1994 (Mellor & Reid, 1994) coined the five domains: 1) Nutrition, 2) Physical environment, 3) Health, 4) Behavioural interactions and 5) Mental state/affects. The 'five domains' do not propose a definition of welfare but aim to provide a scale to allow quantification of welfare compromise within each domain, giving an overall picture of welfare. Whilst the first four freedoms and domains remain broadly the same, the key difference between the two models is the change of the fifth freedom. Changing of the wording from 'freedom from fear and distress' to the 'mental state' domain reflects the importance of considering both positive and negative experiences when assessing overall welfare. In addition, the 'five domain' framework emphasises that the first four broadly physical domains can also contribute to the animals 'mental state' in the final domain. For example, disease may cause poor physical health and influence psychological well-being due to pain or distress resulting from the disease. Since established, the 'five domains' system has been revisited and adapted several times to incorporate current approaches to how scientists approach animal welfare thinking (Mellor, 2017; Mellor et al., 2020). Although developing on the ideas from the 'five freedoms,' the 'five domains' are an additional welfare

consideration rather than a replacement. The 'five domains' may offer an opportunity for a more scientific approach to welfare evaluation, allowing better quantification of welfare state. The 'five freedoms,' however, are still a relevant and a practical consideration when monitoring animal welfare on a day-to-day basis (Webster, 2016).

Measuring animal affective state and emotion is fundamental in getting the best view of overall welfare. However, as emotions are subjective experiences, it is extremely difficult to categorise and quantify them. To study affective processes, there are two conceptual frameworks considered: discrete or dimensional. Focusing on discrete emotions explores the basic emotional systems which have a clear associated brain systems and an adaptive function. For example, experiencing fear may elicit a survival response in a challenging situation. Focusing only on discrete emotions, however, can overlook some more complex emotional states, particularly positive ones. Adapted from what is known about human emotion, a dimensional framework suggests that each emotional state can be quantified along a two, or more, dimensional scale, such as valence and arousal (Kremer et al., 2020; Mendl et al., 2010). Valence reflects perception of the emotion as a positive or negative experience (pleasant or unpleasant) and arousal reflects the activation level of that emotion. For example, excited and content feelings are both positively valenced but arousal would be higher when experiencing excitement (see Figure 1). Subjective experiences characterised in terms of valence and arousal are often coined *core affect*. Michael Mendl and colleagues (Mendl et al., 2010) proposed a framework where both discrete and dimensional approaches are integrated due to the reciprocal interactions. With discrete emotions contributing to the overall position within the core affect space and how this positioning can then influence discrete emotional responses.

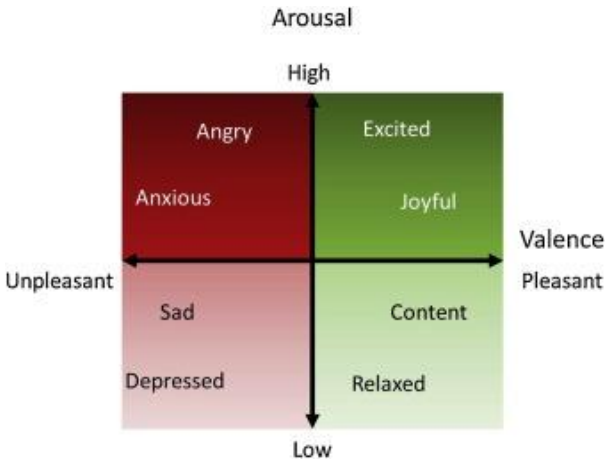

Figure 1: Examples of where discrete emotions fit into the core affect space, in terms of their valence and arousal (Kremer et al., 2020).

1.1.3 Considering the 'good' and the 'bad'

When considering animal welfare, a core component to consider is the presence of positive emotions and experiences, not just the negative effects. The importance of considering both positive and negative influences on welfare is reflected in the AWA (Animal Welfare Act, 2006) in which a 'duty of care' to animal's is to promote positive welfare, not just avoiding negative welfare. However, just as the study of human welfare is often biased by focusing on the negative experiences, so too can animal welfare science. This is generally as negative components are often more intense, meaning they are easier to detect and quantify in comparison to positive experiences (Boissy et al., 2007). However, an understanding of both is vital in order to assess the long-term welfare of animals.

When we considered the long-term welfare or lifetime experience of individual animals, often discussed are the concepts of a 'life worth living' and the animal's 'quality of life.' In humans, this would imply that the individual has a perception about their overall welfare and can weigh up whether the sum of their experiences has given them a good or poor quality of life. In animals, this is a much trickier concept to assess due to the subjectivity of animal experience. In a similar way to welfare assessment, parameters can be used to estimate and judge quality of life in

animals. Mostly, non-human quality of life assessment focuses on the psychological experiences of animals, although some physical parameters can add value too. As quality of life is about the balance between positive and negative experiences, it is crucial to look for indicators of both positive and negative welfare to make an effective judgement. It is also vital to consider how experiences may differ between species and individuals. For example, an animal's past experience can determine how they perceive and cope with different experiences, whether they consider them to be positive or negative and to what extent (Broom, 2014; Green & Mellor, 2011). Furthermore, quality of life is about the overall balance throughout an animal's life, rather than how they feel at any one time. Fluctuation between positive and negative net experience is likely, as different states effect influence at different times. For example, a hungry animal may be consider to have a negative welfare state until they eat and then that experience may have no lasting impact (Mellor, 2016).

1.1.4 Measures of animal welfare

To effectively identify refinements that may improve the lives of animals, it is imperative to be able to assess their welfare. Although there is a lack of 'gold-standard' welfare assessment, there are many measures that can provide information about the biological, affective, and natural state of animals. These measures usually fall into one of two categories. *Resource-based measures* are used to assess the availability of resources to an animal, for example by quantifying the space they are provided, their access to food, water etc. *Animal-based assessment* includes many physiological and behavioural measures, each of which aim to offer insight into how the animal is coping in its current environment; for example, measurements of animal pain and stress (Hubrecht, 2014). Different methods of welfare assessment, both resource and animal-based, may offer insight into one or several of the three factors considered key to overall welfare. For example, by observing an animal's behaviours, it is possible to look out for natural behaviours, clinical signs of pain and behaviours associated with particular affective states.

No single measure is able to provide an overall picture of animal welfare. For this reason, different aspects that contribute to welfare are measured instead. These components can usually be assessed and quantified using either physiological or behavioural (or both) parameters.

Measures that provide information about animal health are clearly beneficial to the animal welfare picture. An animal that is free from pain and disease will undoubtedly have better welfare than one that does not. Pain can elicit a substantial behavioural response in some species, such as the avoidance of use of a painful limb (Broom & Johnson, 1993). However, some species, particularly prey animals, do not display an observable response. In these instances, alternative measures may be applied to identify pain. For example, self-selection analgesia can be used and such studies have shown that animals suspected to be in pain will preferentially select food containing analgesic (Colpaert et al., 1980; Danbury et al., 2000).

A common approach in animal welfare assessment is to ascertain the presence or absence of stress. Defining stress poses a similar problem to that of animal welfare, with many varying definitions and ambiguity about what is involved. Taken from 'Stress and Animal Welfare' (Broom & Johnson, 1993), one definition is that stress is 'an environmental effect on an individual which overtaxes its control systems and reduces fitness or appears likely to do so.' A stress response involves the activation of biological defence mechanisms upon the perception of a threat by the central nervous system. This consists of a combination of behavioural, neuroendocrine, immune, and autonomic nervous system (ANS) responses (Moberg, 2000). The acute stress response is not inherently bad but an adaptive response to enhance animal fitness by, for example, causing a spike in adrenaline that will allow for quick evasion from a predator. Acute stress is often brief and aims to restore homeostasis. However, it can progress into an aversive state, in which coping processes may fail due to severe, sustained, or cumulative exposure to stressors (Moberg, 2000; National Research Council, 2008). Chronic stress has often proven a high-risk factor leading to the deterioration of health in humans and animals, for example leading to increased risk of cancer,

cardiovascular diseases, and immunosuppression (Barnett & Hemsworth, 1990; Hurst et al., 1976). Commonly used indicators of stress and poor welfare include the presence of behavioural stereotypies, a decline in health and fertility, an increase in circulating cytokines, and cortisol hormone (Moberg, 2000).

Focusing on natural behaviour and natural living conditions can be used as a means to assess animal welfare. Written into the five freedoms and considered to be a key component of welfare, natural behaviours are those that an animal will perform under natural or wild conditions (Bracke & Hopster, 2006). For example, providing social contact in primates and grazing opportunities for cows are applied to improve animal welfare. An important distinction must be considered between what is natural for the animal and what they actually want to do. In prey species, it is natural to evade predators but this is clearly not good for welfare (Dawkins, 2006). Understanding what an animal wants is one of the two key questions proposed by Marian Dawkins (the other being 'is the animal healthy?') that need to be asked to obtain the best understanding of animal welfare (Dawkins, 2003). Experiments can be designed such that we are able to ask animals not only what they want but how much they want it. For example, mink are willing to pay the 'cost' of opening heavy doors in order to access a water pool for swimming (Mason et al., 2001).

Whilst we cannot quantify the subjective experience of emotions directly, inferences about emotional states can be made through various physiological and behavioural measures (Neethirajan et al., 2021). Physiological indicators are frequently associated with negative affective states, such as stress. Such indicators may include endocrine changes, such as glucocorticoid release, and the performance of the ANS, e.g. by identifying alterations in heart rate and blood pressure. Although providing value to the understanding of emotions, such measures can be unclear. Different pleasant and unpleasant experiences may elicit the same physiological response. Therefore, while they may provide information about where the emotion

sits within the arousal plane of the core affect space (see Figure 1), they give no indication of the emotional valence (Paul et al., 2005). For example, an increase in heart rate occurs in response to both excitement and fear. Valence may be detectable using behavioural measures, such as simple approach and avoidance behaviours to gauge an animal's perception of a particular stimulus. However, behaviours can still offer a layer of ambiguity, particularly in the absence of context. For example, in distinguishing if an animal is resting or engaging in depressive-like inactivity. Although limited information can be ascertained from a single measure, taken together, multiple measures of affective state can offer great insight into the welfare of animals (Paul et al., 2005).

Whilst many of these measures can help to understand discrete emotions, an animal's underlying emotional state should also be considered, as it can influence how they behave in particular situations. Depressed individuals generally anticipate more negative outcomes and have more pessimistic attitudes. Cognitive or judgement bias tests, which are experiments that work to measure responses to ambiguous stimuli, can be beneficial in helping to identify such an objective experience (Bethell, 2015).

Although there are many potential parameters that may be used to quantify welfare, applying these in practice comes with problems. Measuring multiple welfare parameters is considered good practice to get a broader picture of welfare state but drawing conclusions then becomes difficult if those parameters show contradictory results. Furthermore, many welfare parameters observe adaptive responses to environmental change and the changes in these parameters can be evoked by both positive and negative influences. For example, sexual arousal and predator evasion can lead to similar physiological changes. Similar behavioural responses can be observed in a fearful or relaxed animal when emerging from a shelter, for example with slow emergence indicating reluctance to come out or less erratic movement due to a calmer internal state (Dawkins, 2006; Mason & Mendl, 1993). Additionally, some methods of welfare assessment have the potential to themselves affect well-being. For example, restraining an animal to collect a

blood sample may cause an acute stress response. Care must be taken when selecting appropriate measures and methods, as well as when interpreting results, to avoid causing or detecting welfare changes resulting from the research procedures used. Furthermore, there are lots of factors, such as species, age, and sex that can influence individual coping mechanisms and therefore directly impact welfare. Understanding these factors and animal preferences is a key component to optimising animal welfare. For example, providing a cage mate for an animal would have a very different impact on welfare depending on whether that animal was from a social or solitary species (Broom, 2014; Goymann, 2012). Individual variation in both physiological and behavioural responses is common both between and within species. Upon exposure to a threat, for example, many individuals will flee, some may confront the situation and others may simply freeze.

As a whole, the consensus seems to be that animals can experience emotions and that their affective states contribute to their welfare. However, in order to make improvements to animal welfare, a scientific approach is required (Fraser, 2009; Mellor et al., 2009). To implement changes to the lives of animals, in order to improve their welfare, the requirements within legislation usually want us to provide scientific evidence before we can do so. The subjective experience of emotions means that obtaining hard evidence about affective states is extremely difficult. Therefore, by demanding this evidence, experiences that diminish animal welfare will be allowed to continue. However, whilst this evidence is required the best we can do for animal welfare is to aim to contribute to the knowledge base and further the understanding of animal experiences.

1.2 Animals in research

Whilst human experimentation (with informed consent) does have a role to play in advancing biomedical science, there are many essential experimental practises for which humans cannot be used, as it would be deemed unethical and a violation of their human rights. The use of animal models for scientific research dates to as early as ancient Greece and began to flourish during the Renaissance era, continuing to the modern day. Animal models have been used, and continue, to further our knowledge on anatomy, physiology, and disease. Experiments on living animals in the 17th century led to an understanding of the human heart and circulation system and in the early 20th century testing of new medical treatments on animals, such as penicillin and other antibiotics, has led to the life-saving treatment of multiple human diseases. Almost all medical advances, especially within the last century, have been dependant on the use of animals and of the 111 Nobel prizes awarded for physiology or medicine, around 100 required animal models (Hubrecht, 2014; Understanding Animal Research, 2007).

Whilst emerging non-animal methods may now play an important role in biomedical research, they still cannot fully replace all animals as research models. Although animal research is most associated with improvements to human health, there are several other reasons that animals may be used in research. These may include, for example, safety testing of household products and methods to improve animal health, welfare, or productivity. Annual statistics published by the Home Office provide an overview of the types of research involvement of different animal species. In 2021, 3.06 million procedures were conducted on living animals, of which 1.73 million of these were experimental procedures, the rest being procedures to create and breed genetically modified animals (Home Office, 2022). The purposes of these experimental procedures can be seen in Figure 2. Basic research accounts for the majority of experimental procedures involving animals which includes studies into the function and diseases of, for example, the nervous system. Applied research refers to the development of treatments or

prevention of diseases, such as in cancer research. Regulatory research involves testing the safety and efficacy of products as part of a legal requirement to license their use, for example toxicity testing of human and veterinary medicine.

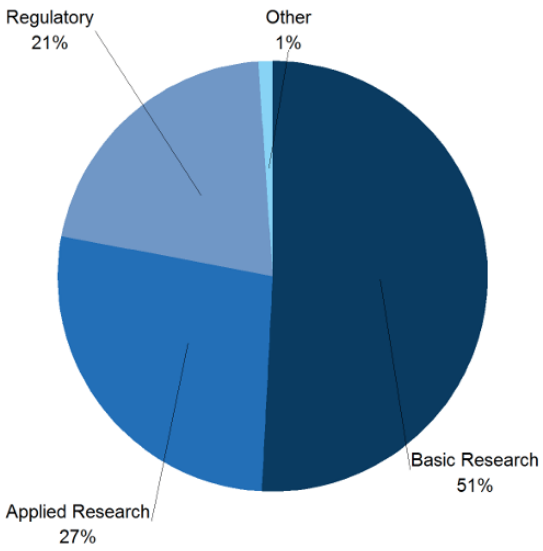

Figure 2: Experimental procedures conducted on living animals, by purpose, in 2021 (Home Office, 2022).

After centuries of being used in research, animal models are still considered vital for the success of future scientific advances. However, as our understanding of animals has developed over this time, the way in which animal experiments are designed and carried out has adapted. The growing knowledge and public concern about whether animal's experience pain and suffering has led to the implementation of legislations to control the ways in which animal's may be used for scientific exploration. The control of animal experiments in the UK began in 1876 when Parliament passed the 'Cruelty to Animals Act' to institute licenses and set limits on animal use. Today, A(SP)A (Animals (Scientific Procedures) Act, 1986) regulates the use of animals for research purposes. It acts to justify animal use by comparing the likely benefits of the scientific outcome with the likely harms to the animals and regulates procedures performed on 'protected animals' with the aim to minimise any pain, suffering, distress, and lasting harm (PSDLH) caused by these experimental protocols (Home Office, 2014b).

Whilst animal models are still considered vital for scientific progress, increasing transparency and raising awareness of the likely benefits can aid public understanding and support (Mendez et al., 2022). The 2014 Concordat on Openness on Animal Research (Understanding Animal Research, 2014) comprises signatures of 128 UK organisations (including the University of Oxford), committing to proactively and clearly communicating their research to the public. Furthermore, the primary goal of organisations such as Understanding Animal Research (UAR) and the European Animal Research Association (EARA) is to offer opportunities for the public to become better informed and learn about animal research.

Public attitudes towards animal use varies and UK opinion polls in 2016 and 2018 revealed a 71% and 68% agreement, respectively, that the use of animal's in research is acceptable providing that 'there is no unnecessary suffering to the animals and there is no alternative' (Ipsos MORI, 2016, 2018). Furthermore, public acceptance rose to 75% agreement with this statement in a survey conducted by UAR during the early stages of the Corona virus pandemic in March 2020 (Williams, 2020). As well as an ethical, moral, and legal obligation to minimise pain and distress in animal experimentation, there is also a benefit to the scientific output. Optimising welfare is likely to ensure the validity of results and allow accurate scientific conclusions to be drawn, as an animal's welfare state can affect its suitability as a research model (Prescott & Lidster, 2017).

The word 'suffering' is frequently mentioned when discussing animals as research models. Suffering can refer to a wide range of subjective emotional states, such as pain, fear, boredom, and hunger (Dawkins, 2008). In broad terms, suffering is likely to occur when an animal is prevented from performing an action which they are highly motivated to do.

In the context of animal research, unnecessary suffering means any pain or distress experienced by the animal that was not required in order to achieve the scientific outcome. Although laws governing animal research state that animals should not suffer unnecessarily, this

does not mean that an animal will not suffer at all. To successfully achieve scientific objectives, the types of experiments required in biomedical research programmes still expose animals to many potential non-natural stressors. Among these stressors are contingent harms, which occur unintentionally as a consequence of animal use. For example, laboratory housing subjects animals to considerable spatial and social constraints, particularly compared to their wild counterparts. Direct harms are those that occurs as a direct result of the experimental procedures themselves and they are an unavoidable consequence of this type of research (Hubrecht, 2014; Russell & Burch, 1959). For example, the use of restraint and potential pain caused by conducting procedures on awake animals, such as blood sample collection. Consideration of both direct and contingent harms is essential to research animal welfare.

As one of the accepted definitions of good animal welfare is adherence to the 'Five Freedoms' (Brambell, 1965), animal research models are particularly vulnerable to suboptimal welfare, as the nature of procedures conducted will impact at least one of these freedoms. For example, food or fluid regulation impacts an animals freedom from hunger and thirst and performance of surgical procedures impacts their freedom from pain, injury, and disease (Honess & Wolfensohn, 2010). Furthermore, the 'Five Freedoms' are cited in the advice notes for conducting harm-benefit analysis (see section 1.2.1), offering a useful framework for the assessment of contingent and direct harms (Home Office, 2015).

1.2.1 The harm-benefit judgement

It is essential that animal research is appropriately justified, such that the likely benefits accrued from the research outweigh the likely harms the animals' experience. The laws that govern animal use in the UK require potential animal experiments to be scrutinised through a harm-benefit judgement (see Figure 3). This involves the weighting of animal suffering against the likely benefit of the research and the overall scientific importance of that research. Various, relevant expertise are used to determine whether the research outcome is worthy of the likely suffering and, where the benefits are deemed worthy, also explores potential refinements to minimise suffering and/or optimise those benefits (Hubrecht, 2014).

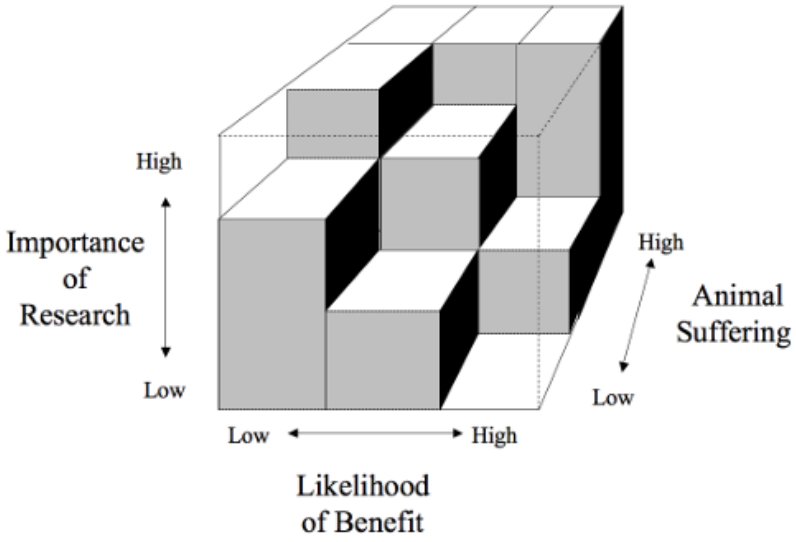

Figure 3: An illustration of the harm-benefit judgement to determine whether a research project should proceed (clear space) or not (solid space) [redrawn by (Pickard, 2013), from (Bateson, 1986) and (Bateson, 2005)].

To help inform this harm-benefit judgement, prospective severity limits are assigned to experimental procedures prior to the granting of animal licenses. Severity limits are classified as 'sub-threshold,' 'mild,' 'moderate,' 'severe,' or 'non-recovery.' In addition, after performing experiments, researchers are required to report the actual severity of those procedures. The actual severity is taken from observations of the animals and is reported for each individual animal as the highest level experienced at any time throughout the experiments, inclusive of cumulative effects. The reporting of actual severity aims to improve transparency and can help to

inform avenues for future refinements (Animals (Scientific Procedures) Act, 1986; Home Office, 2014b).

Selecting the right species for experimental protocols is a key factor in the harm-benefit judgement, where the aim is to keep animal suffering to a minimum whilst maximising scientific output. Despite the uncertainties around definitions and how we interpret the feelings of animals, sentience still plays a key role in the legislative decisions that must be made about which animal species should be protected. As we understand the most about human sentience, humans tend to serve as a 'gold-standard' and certain similarities to humans are used as indicators of sentience in other species. For example, species showing homologies to humans in brain structure and nervous system anatomy are afforded a higher degree of sentience. Other factors such as neural complexity, behaviours and cognition also play a role in the assignment of sentience. Within A(SP)A, these factors have led to the inclusive of all vertebrates (except humans) and cephalopod species as 'protected animals.' Furthermore, some species have additional protection requiring demonstration that no other species is suitable and stipulating that researchers using those animals adhere to additional licence conditions. These species include non-human primates (NHPs), dogs, cats, and horses. Only 1% of UK scientific procedures were carried out using these specially protected species in 2021 (Home Office, 2022). However, the use of these animals still causes wide public concern.

1.2.2 Requirements for conducting animal research

There are several requirements set out in the A(SP)A (Animals (Scientific Procedures) Act, 1986) that must be met before scientific experiments on any protected species can be performed. Firstly, three licences from the Home Office are required. These include an establishment licence, associated with the place in which the work will be carried out and a project licence (PPL) linked to the programme of work. Each individual working under the PPL, carrying out procedures, must also hold a personal licence (PIL). Establishments conducting animal research are legally obliged

to establish and maintain Animal Welfare and Ethical Review Bodies (AWERB). AWERB play a key role in promoting the awareness of animal welfare. Amongst other tasks, AWERBs promote the 3Rs (see section 1.2.3), support staff and can provide advice to the establishment licence holder about whether to support project proposals. AWERBs are actively involved throughout the entirety of a PPL, following and reviewing the development and outcomes of the scientific procedures.

To help to ensure compliance with the terms of PPLs and PILs, establishments are subject to regular Home Office inspections. Furthermore, establishment licence holders must appoint named persons who should help them to fulfil their responsibilities. These include a 'Named person responsible for compliance,' a 'Named Animal Care and Welfare Officer,' a 'Named Information Officer,' a 'Named Training and Competency Officer,' and a 'Named Veterinary Surgeon.' Their roles include ensuring licence conditions are complied with, overseeing animal care and welfare, ensuring access to information, making sure staff are appropriately trained and advising on animal health and treatment (Animals (Scientific Procedures) Act, 1986; Home Office, 2014b).

1.2.3 The 3Rs: Replacement, reduction and refinement

An integral part of legislation governing animal research in the UK (Animals (Scientific Procedures) Act, 1986) is the 3Rs, serving as a framework to promote the production of high quality and humane animal research. The 3Rs concept was proposed by Russell and Burch in 1959 as a method for approaching animal experimentation to minimise suffering and improve animal welfare. In order to achieve this, Russell and Burch outlined three key principles: replacement, reduction and refinement (Russell & Burch, 1959). Replacement aims to remove sentient animals from potentially distressing experiments by replacing them with non-sentient alternatives, such as computer models or microorganisms. The reduction principle also seeks to remove animals from the research environment by ensuring that the minimum number of animals is used for

experiments, whilst still maximising the scientific output (Tannenbaum & Bennett, 2015). When the use of animals is deemed unavoidable and necessary, the refinement principle works to improve the welfare of the animals involved. For example, in husbandry practices, providing species appropriate housing to meet animal needs and, in experimental practices, decreasing the incidences or severity of scientific procedures. Consideration of the 3Rs is essential for conducting animal experiments in the UK, with scientific organisations dedicated to highlighting, promoting, and encouraging these principles. For example, the UK-based organisation the National Centre for the Replacement, Reduction and Refinement of Animals in Research (NC3Rs). Each of the 3Rs offers opportunity to reduce animal suffering and, legally, they must always be considered when planning and undertaking animal experimentation. Furthermore, evidence of implementing the 3Rs is often a requirement by funders when choosing to support research and compliance with the 3Rs is monitored by an AWERB when establishing support for project licenses. Scientific investigation of the 3Rs is continuously providing further evidence of alternative approaches that will improve the lives of animals. A few examples of such studies include refinements in mouse handling techniques that will reduce stress (Davies et al., 2022), reducing animal use by using repeated within-animal measures (Walther et al., 2023) and seeking *in vitro* alternatives that will completely replace particular *in vivo* experiments (Chichagova et al., 2023).

Implementation of the 3Rs is not always straightforward. Opinions and priorities differ between individuals, and this can cause conflict between the 3Rs. For example, when trying to balance reduction and refinement, where fewer animals may be used for an experiment but at the cost of a higher degree of suffering. Additionally, animal researchers may have concerns about the impact of the 3Rs on their experimental outcomes and may require further evidence to have confidence enough to implement them (Hubrecht, 2014).

To successfully implement the 3Rs, particularly refinement, and minimise suffering, it is crucially important to be able to assess the welfare of animals. This will allow for the

quantification of the animal's experience, highlighting the impact of specific scientific procedures and husbandry conditions on physical and psychological wellbeing. In the planning stages of animal research, using information from previous welfare assessments allows for more well-informed decisions to be made when conducting the harm-benefit analysis. Furthermore, monitoring animal welfare and experiences may highlight potential areas for refinements that may further minimise suffering and can provide insight for the planning of future research.

1.2.4 Cumulative experience

Animals used in biomedical research experience multiple invasive procedures and contingent stressors arising from laboratory housing. Whilst some of these will have an effect on acute welfare, it is also imperative to understand whether these experiences having a lasting, cumulative effect. Frequent repetition and continual exposure to stressors (sometimes with multiple stressors occurring simultaneously), emphasises a need for cumulative effects to be monitored in laboratory species in order to gain a full view of animal welfare. In fact, legislation governing the use of animals in research requires consideration and assessment of this cumulative effect (Animals (Scientific Procedures) Act, 1986).

Simply adding together all of the stresses or harms that research animal has been exposed to does little to understand their cumulative experience (Hubrecht, 2014). Factors such as individual variation, underlying mental state, and occurrence in proximity to other stressors can hugely influence how a stressful situation is perceived by an animal. Furthermore, repetition of stressors means that animals may know what to expect next time and this may result in habituation to the repeated procedure or hyper-sensitivity to it (Pickard, 2013). The following definition of cumulative experience attempts to encapsulate all the key elements to be considered: 'the sum of all events and effects, including their quantity, intensity, duration, recovery between and memory thereof, that impact, adversely, positively and by the way of amelioration on the welfare of an animal over its lifetime' (Pickard, 2013). The use of the word

'positively' in this definition is an essential component for true consideration of cumulative experience. It allows for both positive and negative experiences to be accounted for, and thus creates a better picture about the animal's quality of life (Green & Mellor, 2011).

Repeated exposure to laboratory stressors may have additive or non-additive effects. A non-additive effect would imply that, whilst certain procedures and events may impact welfare acutely, there is no lasting effect and there is full recovery to the baseline welfare state. An individual may even habituate to an aversive event, thus reducing the acute welfare impact of that event upon each subsequent exposure. Additive effects, however, highlight a potential residual effect, leading to a 'stacking up' of diminished welfare, which may eventually impede the significance of experimental results. This will occur when an experimental protocol has a long lasting impact or if there is an insufficient recovery period before the repetition of that (or another) protocol (Pickard, 2013). Identification of situations and procedures that have this additive effect is essential, as it can highlight areas for improvement and force implementation of changes that benefit overall welfare. For example, ensuring that sufficient recovery time is provided between events to avoid accumulation of stress. It is also essential that measures of cumulative experience should take into consideration both positive and negative experiences, in order to best create a picture of an animal's quality of life.

1.3 Non-human primates as research models

The use of non-human primates (NHPs) is still considered essential for many areas of biomedical research. Sharing a close phylogenetic relationship and many homologous features with humans, makes them the best available animal models for many scientific questions (Carlsson et al., 2004; Phillips et al., 2014; Weatherall, 2006). Legislation in the UK (Animals (Scientific Procedures) Act, 1986), currently permits the use of NHP species for research purposes. However, NHP are classified as a 'specially protected species' and researchers seeking to use primates in their experiments must provide additional evidence that no other species are suitable

for the purpose of that research (Bateson et al., 2011). NHP models can be used in a variety of study types, such as fundamental biological research and the development and testing of human medicine. A 2004 report showed that the most common areas of research for NHP models are microbiology, neuroscience, biochemistry, and pharmacology (Carlsson et al., 2004). Furthermore, a 2017 EU SCHEER report highlighted areas of fundamental, translational, and applied research for which NHPs continue to be used. These included the development of pharmaceuticals, treatment of infectious disease and neuroscience, among others. The report discusses currently available alternative methods to NHP models and highlights an ongoing need for NHPs to continue scientific progress in these fields (SCHEER, 2017).

Neuroscience is one of the key research areas for which NHP use is especially valuable. Due to their common ancestry, NHPs share many physiological and psychological similarities to humans, such as complex cognitive capabilities, socialisation abilities, and similarities in the structures of their central nervous systems and neural circuit organisation (Capitanio & Emborg, 2008). NHP brains, compared to other mammals, closely resemble human brains in many ways. For example, in prefrontal cortex size, degree of myelination, and overall brain size (relative to body mass) (Phillips et al., 2014). Their level of intelligence and ability to perform fine motor skills mean that NHPs can be trained to carry out complex tasks. Performance of these tasks can be vital for understanding some of the brains mechanisms of cognitive functioning (Roelfsema & Treue, 2014). Furthermore, many NHPs, particularly Old world monkeys, also share aspects of their lifestyles (e.g. diurnal, omnivorous), sensory and anatomical abilities (e.g. opposable thumbs) with humans (Phillips et al., 2014). These similarities have led NHPs to play a fundamental role in basic and translational neuroscience research. Using NHP models has developed our understanding of and provided treatment options for many neurological diseases, such as Alzheimer's and Parkinson's. A continued need for NHPs is emphasised by the neuroscience community, to achieve a fundamental understanding of the human brain and to

develop effective treatments for neurodegenerative and neuropsychiatric conditions (Mitchell et al., 2018).

The characteristics of NHPs that make them valuable research models, however, are also the same characteristics that indicate a high degree of sentience and suggest that NHPs are likely to experience pain and suffering (Perretta, 2009; Rennie & Buchanan-Smith, 2006a). For example, their neurophysiological complexities offer the potential for experiences of pain and pleasure to be similar to humans and offer the ability to anticipate and recall these experiences, thus enhancing them (Prescott, 2010).

Although support among the British public for animal research was at 68% in 2018, support for research using NHPs was much lower (12-16% showing support for medical research using any NHP species) (Ipsos MORI, 2018). This implies that although generally supportive of animal research, asking the public to consider research by species may trigger an emotional response making it harder to accept the use of certain animals. In the case of NHPs, an animal group which the public are least supportive of, it could be their close relation and human-like qualities that evoke this emotional response (Mitchell et al., 2021). Further evidence of the influence of human similarities to the public perceptions of research in certain species is reflected in the use of great apes. Fewer countries now permit the use of great ape species in research, with no great apes used in the UK since 1986 and their use banned in the UK in 1997 (Hubrecht, 2014; Weatherall, 2006). Interestingly, as an increased support for animal research overall was seen during the early stages of the COVID pandemic, so to was increased support for NHP use. 73% of those surveyed said that it was acceptable to use NHPs (amongst other species) to develop tests, treatments, and vaccines for COVID-19, if it was the only way to achieve this (Williams, 2020). Furthermore, openness from researchers about their research using NHP can increase public support and understanding (Mendez et al., 2022).

Rhesus macaques (*Macaca mulatta*) are one of the mostly commonly used NHP species in biomedical research and the most used in studies of neuroscience (Carlsson et al., 2004). Rhesus macaques share 93% of their genes with humans (Gibbs et al., 2007), which is higher than most other animal models, such as mice (85%) (National Human Genome Research Institute, 2010). This in addition to many other human homologies, such as similar life stages, brain structures, and nervous systems make them valuable research models. They are also a non-endangered primate species, who are relatively adaptable and they tend to cope well within the laboratory environment (Bernacky et al., 2002). Furthermore, their ability to perform complex cognitive tests is an essential requirement for many neuroscientific studies (Stonebarger et al., 2021).

The typical life of a rhesus macaque used in neuroscience research involves regular exposure to potential stressors. Common, often daily, stressors may include movement restraint (e.g. primate chair), food or fluid control to boost motivation, and social separation during testing sessions (Pfefferle et al., 2018). In addition, NHPs can be exposed to more experimental procedures, such as the administration of general anaesthesia (GA) for the purpose of collecting brain images or to conduct a surgical procedure. Exposure to such stressors may have a marked impact on animal welfare, particularly if exposure is prolonged or repetitive.

Macaques used in neuroscience research may experience compromised welfare caused by living in a captive environment, especially when that environment differs significantly from the wild environment for which they have evolved (Prescott, 2010). Wild rhesus macaques live in large multimale-multifemale groups (10-50 members) and can range freely over large areas, with average ranges of 1-2kms per day (Fooden, 2000; Magden et al., 2015). Even though the UK's minimum housing size for macaques is one of the most generous in comparison to other country standards, it is still only a fraction of their wild range. Current housing standards were established by considering NHP psychological well-being, aiming to allow NHPs opportunities to engage in natural behaviours. However, housing must also be practical for human users, to conduct

research and care for the animals safely and effectively. Legislation promotes the social housing of macaques used in research, as there is strong evidence that social isolation causes psychological distress and the provision of social opportunities is believed to be the most effective form of NHP enrichment (Rennie & Buchanan-Smith, 2006b). However, stable social housing is not always possible. Animals may end up being subjected to regular social disruption due to compatibility issues with other NHPs available (as there are limited available options within a laboratory environment). When animals are socially incompatible, due to restrictions of space, those issues can become a detriment to physical and mental health. This is caused by an inability to avoid social groupmates, which may promote hypervigilance and sometimes results in fight injuries. In some cases, individuals will end up being singly housed due to a lack of available, compatible social partners or occasionally, when required by the research protocols. This may leave some macaques with insufficient socialisation opportunities.

Neuroscience research often requires conducting complicated experimental procedures that require extensive animal training and so this field commonly requires macaques to be placed in studies for long periods, usually several years (Pickard, 2013). Thus, it is imperative to consider their cumulative experience. Some measures, such as telomere attrition and grey matter volume in the hippocampus, are showing promise as objective indicators of long-term experience (Bateson & Poirier, 2019). Tool such as the 'Extended welfare assessment grid' can provide a useful estimate of cumulative suffering by summarising welfare parameters over time and considering change in those parameters subject to further stressors or the implementation of refinements (Honess & Wolfensohn, 2010). When discussing the assessment of cumulative experience, a recommendation in a review by the Animals in Science Committee (ASC) stated that 'researchers should explore if this can be done by obtaining valid data from animals already undergoing regulated procedures, to avoid causing additional harms' (Animals in Science Committee, 2017). This suggests that the collection of longitudinal datasets from various welfare

parameters, collected alongside existing neuroscience protocols, may offer insight into cumulative macaque experience.

As with all research animals, consideration of the principles of the 3Rs is an essential part of conducting neuroscience experiments on NHPs. Advancements to all three R's have been made in recent years. For example, considering the use of advanced brain imaging techniques (to replace), optimising data yield through the use of multi-electrode arrays (to reduce) and the application of positive reinforcement training (to refine) (Prescott et al., 2017; SCHEER, 2017). However, there are still several barriers to 3Rs implementation amongst the primate community. Key amongst these are a lack of awareness of 3Rs advancements and a reluctance to change from the individuals working directly with NHPs. Efforts must be made to seek out potential refinements when conducting research on animals and, when 3Rs advances are made, to properly validate and test them to encourage their uptake (Lemon, 2018; Prescott et al., 2017). More active involvement from NHP researchers may help with such improvements. For example, by playing a proactive role in raising awareness of new techniques and refinements or in the collection of 3Rs relevant data alongside their own studies.

1.4 Assessing the welfare of non-human primates

Whilst macaques are considered essential models for neuroscience research, striving to improve their welfare is essential. By engaging in research using NHPs, scientists are in agreement that there is no alternative and that there will be no unnecessary suffering for those animals. To avoid unnecessary suffering and to fully adhere to the legal requirement for 3Rs implementation, welfare assessment is crucial for the identification of areas for improvement. In addition to being beneficial for the animals, optimising welfare can lead to better quality science. Compromised welfare makes animals poor research models, leading to unreliable and variable results (Prescott & Lidster, 2017).

Experimental procedures and husbandry conditions both have the potential to cause PSDLH and this is particularly concerning in NHPs due to their perceived higher levels of sentience (Perretta, 2009). Quantifying the impact of such stressors allows for the identification and implementation of refinements, which may improve welfare, benefitting both the animals and the scientific research. Ongoing welfare assessments are therefore an essential component of animal research programmes, ensuring commitment to the 3Rs, as required in the UK by A(SP)A (Animals (Scientific Procedures) Act, 1986; Home Office, 2014b).

Although employing novel techniques, this thesis uses a few, well established behavioural and physiological parameters of welfare assessment to form a picture of the impact of neuroscientific careers on rhesus macaques.

1.4.1 Behavioural monitoring

Behavioural assessment is a vital contributor to the understanding of an animal's welfare. An animal's behaviour reflects its own intentions and can be influenced by many complex processes. Behavioural monitoring has the potential to provide insight into each of the three components of animal welfare: biological function, natural state, and affective state (Fraser, 2008; Watters et al., 2021).

Under the most basic definition of animal welfare, an animal's biological needs should be met, and they should be free from pain and disease. There are several behavioural indicators that can provide information about the health of animals. For example, abnormal or reduced movement may indicate a detriment to physical well-being (e.g. lameness). Hence, behavioural observations form a key component of veterinary diagnosis (Barnett & Hemsworth, 1990).

The presence of natural behaviours (i.e. those behaviours expected of an animal's wild counterpart) can highlight the natural living state of captive animals. Generally speaking, it is considered that animals with good welfare will be engaging in their species' natural behaviours. In

fact, attempts are usually made to encourage such behaviours from captive species, such as the provision of forage and climbing opportunities to laboratory-housed NHPs (Home Office, 2014a). The interpretation of the welfare impact implied by naturalistic behaviours can be difficult in captive environments, where the usual driver for such a behaviour is absent. For example, captive predators have no need to hunt for prey. However, the removal of this behavioural opportunity could be at the benefit or detriment of the animal depending on whether they have an innate desire to perform such a behaviour or not. The welfare of animals may not be dependent on them performing a full repertoire of species-specific behaviours but instead animals should have the capacity to respond in a species-typical way to stimuli within their environment (Rennie & Buchanan-Smith, 2006b).

Incorporation of behavioural measures into welfare assessment can offer insight into affective states. Methodologies such as preference testing can aid in our understanding of what animals want, which is a key component of welfare assessment (Dawkins, 2003). Furthermore, both positive and negative affective states can be explored using behavioural indicators. Whilst some behaviours are associated with negative emotional states, such as anxiety and boredom, others are likely highlighting better welfare, such as the presence of play behaviours or the anticipation of engagement in a rewarding activity (Mellor, 2015; Watters et al., 2021).

When considering the welfare of laboratory housed animals, behavioural signs of poor welfare are often at the forefront of the investigation. Usually occurring within a sub-optimal environment, abnormal behaviours and behavioural indicators of stress play a key role (Hubrecht, 2014). Such behaviours may include stereotypies, self-injurious behaviours, or avoidance behaviours. In the case of social species, such as NHP, social behaviours can also reflect welfare state. For example, increased aggression or fear towards conspecifics could indicate reduced welfare (NC3Rs, 2015). Furthermore, a lack of responsiveness to stimuli is characteristic of depressive-like behaviours and may include a decrease in activity levels (Camus et al., 2014).

Consideration of factors that will contribute to poor welfare is critical in laboratory animal research, helping to quantify their experiences. This is critical for understanding intervention points, retrospective reviews, and for providing insight into the 'cost' of experiments.

The absence of behavioural indicators of poor welfare does not conclusively demonstrate that animal welfare is good. Attention should also be paid to indicators of positive experiences and good welfare state. Engagement in a range of different behaviours, referred to as behavioural diversity, can act as an indicator of positive welfare (Miller et al., 2020). Furthermore, better welfare can be associated with the presence of species normative behaviours (Hubrecht, 2014). Although caution must be taken to ensure that these behaviours are something which the animal wants and is motivated to perform. Comparisons to the activity budgets of their wild conspecifics has been used as an indicator of enhanced welfare in some captive species. However, this assessment may not be so straightforward. For example, activity budgets among wild individuals may vary dependent on multiple factors such as climate, group composition, and geographical range (Howell & Cheyne, 2019).

Within a laboratory environment, regular behavioural monitoring usually forms a key component of the ongoing welfare assessment. Daily observations of research animals is often a regulatory requirement and forms part of the standard operation of most research facilities, providing a crucial opportunity for early identification of health and husbandry issues (Hubrecht, 2014). More frequent observations will be required for effective monitoring following particularly stressful procedures. For example, post-operative monitoring should be conducted at 1-2 hour intervals to adequately assess recovery and be able to quickly identify adverse effects (Flecknell & Roughan, 2004).

Regular monitoring provides a platform for within-individual comparisons of behaviour. There is considerable diversity between individuals in terms of their behavioural responses (Broom & Johnson, 1993). Therefore, intra-individual comparisons can be extremely valuable for

monitoring behavioural change over-time or in response to a particular stressor. Whilst changes to behaviour may reflect change to welfare state, this cannot be assumed. Behaviour is a tool for managing environmental variation and thus changes should be expected in response to particular challenges (Watters et al., 2021).

Obtaining measures of animal behaviour is a relatively simple process, as this is usually achieved through observation which can be done non-invasively and with minimal disruption to the animal. However, experience and appropriate training may be required in order to obtain a reliable measure (Hubrecht, 2014). Although subjective behavioural assessments have their value, a structured system will supply a consistent, quantitative measure. For example, an objective behaviour-based scoring system was more effective at detecting pain in rats than a subjective approach (Roughan & Flecknell, 2006).

An objective, quantitative approach to behavioural assessment may offer great value to the investigation of animal preferences, aversions, and priorities. The ability to collect behavioural data remotely and automatically will be hugely beneficial for the ongoing assessment of laboratory housed NHPs. Automated image analysis makes the detection of behavioural patterns from video recordings a possibility for behavioural analysis (Rushen et al., 2012). However, this technique may not be appropriate for all settings. In the context of NHPs in the laboratory, camera placement can pose a challenge to ensure full visualisation whilst also remaining out of the grasp of NHPs, in order to avoid injury to the animal or damage to the camera. These issues may be overcome through the use of animal-attached devices instead.

Wearable sensors can be used to monitor animal movement and include devices such as GPS trackers, RFID tags and accelerometers (Diana et al., 2021; Rushen et al., 2012). Such devices are usually self-contained, with their own battery and memory storage. This creates opportunity for continuous data collection over extended periods without the confounds of animal disruption

or visibility. Additionally, wearable devices are ideally suited for monitoring group-living animals at an individual level as no identification methods would be required.

Accelerometers are spring-like, piezoelectric sensor that deform to create a signal in response to gravity and movement (Brown et al., 2013). This signal is proportional to the acceleration of the wearer and provides a quantitative estimate of the wearer's movement intensity and duration (Martin & Bateson, 2007; Rushen et al., 2012). Accelerometers can be affixed to animals, for example attached to a neck collar, to provide an alternative platform for exploring animal activity and behaviour.

1.4.2 Glucocorticoids in welfare assessment

Quantifying animal stress is a key element of welfare assessment and will provide insight into the adaptive response of animals to challenging situations. The hypothalamic-pituitary-adrenal (HPA) axis is responsible for the neuroendocrine response to stressors. The HPA axis is activated when an external stimulus is perceived as a stressor and results in a cascade of hormonal responses, firstly from the hypothalamus, where corticotropin-releasing hormone (CRH) is released, which then stimulates the pituitary to produce adrenocorticotropic hormone (ACTH). ACTH stimulates glucocorticoid (GC) release from the adrenal glands into the bloodstream where they will target tissues and organs to re-establish homeostasis, which is the ultimate goal of this axis (Matteri, 2000). Therefore, the activation of this axis can be detected by evaluating GC hormone levels in biological samples (Mostl & Palme, 2002; Novak et al., 2013). In times of acute stress, GC concentrations will peak to drive the redirection of energy from long-term functions to those required to manage the stressor (Dulude-de Broin et al., 2019). This mechanism allows animals to rapidly cope with potential threats, for example, by prompting behavioural change or triggering essential physiological changes, such as an increase in heart rate or respiration (Barnett & Hemsworth, 1990).

In acutely stressful situations, the release of GC operates as a negative feedback mechanism. GC release will reduce the activity of the HPA axis, restoring those essential biological mechanisms from which attention was diverted in order to cope. However, if a stressor is severe, sustained, or sometimes when exposure is repetitive, the HPA axis remains active, prolonging the elevation of the GC concentration (Sheriff et al., 2011). This chronic stress has potentially detrimental effects on animal welfare. For example, decreasing the immune response, reducing growth, and putting individuals at higher risk of developing pathologies (Barnett & Hemsworth, 1990; Moberg, 2000; Mostl & Palme, 2002). Monitoring HPA axis activity can offer insight into an individual's stress coping abilities and can potentially highlight unmanageable stressors,

ultimately resulting in a detriment to animal wellbeing. Therefore, detection of GC levels in various sample materials is a valuable technique which could contribute to welfare research.

The mechanism and functioning of this stress response is highly conserved across the vertebrate species. However, this does not mean that assessment of the HPA axis is always the most appropriate measure and there are numerous confounds that suggest results should be interpreted with caution. Autonomic responses to both pleasurable and unpleasant experiences can be very similar making the valence of these responses difficult to interpret (Dawkins, 2006). Extrinsic factors, such as time of day, season, and social status, and some intrinsic factors, such as age and sex, appear to influence the stress response (Sheriff et al., 2011). Contradictory results have been reported in cortisol response too, with acute challenges being associated with both an increase and decrease in GC response (Mason & Mendl, 1993). There is also a degree of variation between individuals in their responses to the same stress induction. Furthermore, there are controversial thoughts about the magnitude of the HPA axis response. Some state that it correlates to the perceived severity or intensity of the stressor (Broom & Johnson, 1993), whereas other sources believe this information is limited (Otovic & Hutchinson, 2015).

In spite of these caveats, assessing GC levels can still provide a valuable contribution to the overall picture of an animal's welfare. Results just need to be interpreted with care, bearing context in mind and ideally gathered alongside other measures of welfare.

1.4.2.1 GCs in blood

The primary GC produced during stress response differs between species. For example, corticosterone is usually the primary GC in rodents and cortisol is the principal GC in humans and most other mammal species, including primates (Matteri, 2000; Reeder & Kramer, 2005). The most direct and widely used method of cortisol assessment is by obtaining a peripheral blood sample. In most species, changes in blood cortisol (in response to a stressor) can be detected after just 2-5 minutes, with a peak response reached by 30 minutes post-stressor and basal

cortisol levels re-established within 90 minutes (Broom & Johnson, 1993; Mormede et al., 2007; Sheriff et al., 2011). This makes blood cortisol an ideal medium for assessing the immediate stress response. Cortisol levels in the blood stream are frequently used throughout the animal welfare field to evaluate the impact of various physiological and psychological stressors. For example, in cows, plasma cortisol increases were observed after 2 minutes in restraint (Stilwell et al., 2008) and were significantly increased during exsanguination after slaughter, with values twice as high in the absence of prior stunning (Barrasso et al., 2020). Dogs in sheltered housing have elevated blood cortisol levels, particularly during the first 3 days of arrival (Hennessy et al., 1997), and human interaction with dogs work to alleviate stress, with decreased plasma cortisol levels observed (Shiverdecker et al., 2013). Laboratory-housed primate welfare is frequently studied through cortisol assessment. The typical experiences of laboratory life, such as restraint (Morrow-Tesch et al., 1993), housing type (indoor or outdoor) (Schapiro et al., 1993) and social rank (Czoty et al., 2009), exert influence on blood cortisol levels. In one study, high stress-induced plasma cortisol concentrations in the early life of rhesus macaques has been associated with less sensitive parenting behaviour later on (Wood et al., 2021).

Whilst the assessment of plasma cortisol concentration offers the most direct look at HPA axis activity, collection of blood samples is not always possible or appropriate in animal welfare studies. Baseline plasma levels of GCs are hugely variable, particularly as levels naturally fluctuate throughout the day with the circadian rhythm (Otovic & Hutchinson, 2015). The process of sample collection is invasive and can itself act as a stressor, usually requiring the animal to be physically or chemically restrained. This can be counterintuitive to the purpose of the study, where the overall objective is usually to improve animal welfare. In studies conducted on research animals in the UK, according to A(SP)A (Animals (Scientific Procedures) Act, 1986), collection of a blood sample would be classified as a 'regulated procedure', meaning it has the potential to cause the subject a level of PSDLH. Licenses (establishment, project and personal licenses) and personnel training are required to obtain the blood sample needed for cortisol

detection. Furthermore, results could be confounded by a stress-induced response to the collection method, as the activation of the HPA axis causes a change in blood cortisol concentration within minutes (Mormede et al., 2007). For example, a study conducted on blue tits (*Parus caeruleus*) assessed their reaction to capture and handling, reporting a noticeable increase in circulating GC levels after just 3 minutes (Mueller et al., 2006).

Following its release into the blood stream, GCs are found in the form of their metabolites within other biological samples. Detecting cortisol in faeces, urine, saliva, and hair provides alternative, less invasive opportunities to assess HPA axis activity and monitor the stress response.

1.4.2.2 GCs in Saliva

GCs are lipid-soluble steroids, meaning unbound GCs can pass freely from the capillaries into the salivary gland and diffuse into the saliva. There is a close correlation between salivary GC and unbound blood GC levels, although this varies between species (Kirschbaum & Hellhammer, 1994; Sheriff et al., 2011). Therefore, salivary analysis is a valid, comparable, and non-invasive alternative to monitoring stress.

Saliva is most similar to plasma cortisol, as it represents a single time point and cortisol levels fluctuate throughout the day in line with the circadian rhythm, with levels typically peaking in the morning (in diurnal animals) (Otovic & Hutchinson, 2015). There are multiple methods for collecting saliva from animals but most of these take at least 30 seconds, up to a few minutes, to complete. This means that there is some reliance on cooperation from the subjects and a degree of animal training is usually required to obtain the sample. This makes salivary cortisol analysis a more feasible endeavour in captive animals and more difficult in freely moving, wild animals (Sheriff et al., 2011). Thus, saliva collection for cortisol analysis is a well-established tool for welfare assessment in many species (d'Angelo et al., 2021; Menargues et al., 2008), including

multiple primate species (Ash et al., 2018; Pearson et al., 2008) and rhesus macaques (Boyce et al., 1995; Higham et al., 2010; Lutz et al., 2000; Pfefferle et al., 2018).

1.4.2.3 GCs in Urine & Faeces

Metabolites of GCs are detectable in animal excreta and the collection of urine and faeces offers another opportunity for non-invasive monitoring of HPA axis activity. In most species, GC metabolites are measureable in both urine and faecal samples, with urine usually being the main route of GC excretion (Mormede et al., 2007; Sheriff et al., 2011).

Whilst cortisol in plasma samples reveals the immediate stress response, the level of urine and faecal GC metabolites (UGM and FGM) reflect an accumulation of cortisol secretion over time. This makes the results less susceptible to brief cortisol fluctuations, such as those caused by circadian patterns (Mormede et al., 2007; Sheriff et al., 2011). The time period of this accumulation does vary between species but generally, urinary cortisol accrues over a few hours and faecal cortisol over 1-2 days, associated with the gut-passage time of that species. For example, the peak of radiolabelled cortisol excretion in a long-tailed macaques was after 5.5 hours in urine and 26.2 hours in faeces (Bahr et al., 2000).

Both UGM and FGM measures have been used for animal welfare studies in numerous species. For example, UGM has been used to explore the welfare effect of kennelling in dogs (Stephen & Ledger, 2006), housing in pregnant sows (Pol et al., 2002), and the presence of immature offspring in squirrel monkeys (Soltis et al., 2003). FGM assessment is very popular among the animal welfare research community. This is most likely because collection of faecal samples can be done without any disruption to or expectation from the animals being sampled. Many studies on both wild and captive animals have used FGM to document stress, (Ganswindt et al., 2003; Hadinger et al., 2015; Rauch et al., 2014), including studies from macaque species (Marty et al., 2017; Ostner et al., 2008; Young et al., 2014).

Although urine and faecal sampling is a simple and effective method to track HPA axis activity, some caution is needed. There are numerous factors that may influence cortisol concentrations and lead to misinterpretation of results. For instance, diet may affect gut transit time and GC levels in faces (Otovic & Hutchinson, 2015). Urinary cortisol concentrations may not be accurate if the whole urine sample is not analysed (Sheriff et al., 2011). Furthermore, avoiding contamination is essential, else results will be skewed by cortisol present in the contaminant. For example, assaying FGM from a sample saturated by urine would give a falsely inflated GC concentration as GC levels in urine are much higher than that found in faeces (Mostl & Palme, 2002).

1.4.2.4 GCs in Hair

Chronic stress can refer to both an animals' constant, underlying state and the accumulation of successive exposure to acute stressors (Ladewig, 2000). For example, research animals are permanently living within an 'unnatural' environment and are continually exposed to stressful events, e.g. human presence, handling, experiments etc. Long-term stress is still not well understood in animals but its assessment is crucial for a full understanding of welfare state. Whilst all of the aforementioned mediums of GC analysis give insight to the influence of transient stressors, substantial repeated sampling from the same subject would be vital to explore the level of chronic stress. Detection of cortisol in the hair can provide an opportunity to explore the long-term activity of the HPA axis and thus, offer a measure of chronic stress (Davenport et al., 2006; Meyer & Novak, 2012; Sheriff et al., 2011).

The exact mechanism of cortisol incorporation into the hair is not fully understood. The major integration is believed to be via the capillaries that supply blood to the hair follicle and so, cortisol from the bloodstream becomes incorporated within the hair shaft as it grows (Davenport et al., 2006; Sheriff et al., 2011). Therefore, analysis of cortisol in hair strands can provide a 'life-history' of HPA axis activity during its growth and segmental analysis of a hair strand could

provide a retrospective calendar of cortisol release (Carlitz et al., 2014). However, there is a time delay which must be considered between the cortisol incorporation into the hair and the time from which it can be measured. This is due to cortisol entering the hair during the anagen (active) phase of hair growth, during which time the hair strand has not arrived at the surface of the skin (Heimburge et al., 2019). Besides via the bloodstream, there are other potential routes for cortisol to get in the hair follicle; including secretions from the subject, such as sebum and sweat, and contamination from external substances, such as faeces, saliva, and urine. These additional sources could potentially confound results (Koren et al., 2019; Meyer & Novak, 2012). Although it is unknown whether these secretions and contaminants can become incorporated into the hair follicle, usually the hair is washed prior to assay in order to remove the deposits from the surface.

In 2006, cortisol concentration in the hair of rhesus macaques was shown to provide an index of HPA axis activity during hair growth and highlighted its usefulness as a chronic stress measure (Davenport et al., 2006). Following this, detecting cortisol in hair samples from animals has been employed as a method for long-term stress assessment in many species (e.g. rabbits (Peric et al., 2018), goats (Dulude-de Broin et al., 2019), great apes (Carlitz et al., 2015; Carlitz et al., 2014), pigs and cattle (Casal et al., 2017; Heimburge et al., 2020)). Furthermore, more studies from rhesus macaques have since used hair cortisol as a measure of chronic stress in relation to social status (Vandeleest et al., 2019), population density (Dettmer et al., 2014), depressive-like behaviours (Qin et al., 2015), hair loss, alopecia and pregnancy (Lutz et al., 2019; Novak et al., 2014).

For researchers, assessment of hair cortisol offers additional benefits. The collection method is relatively simple and non-invasive, although may require some cooperation from the animal either via training or restraint. Additionally, cortisol within hair samples offers good stability. This allows samples to be stored for an elongated period of time, at room temperature. For example, one study assessed HCC in samples taken from a museum specimen polar bear (>80

years old) and found that HCC were similar to samples taken from recent bears, showing no evidence of cortisol degradation over time (Bechshoft et al., 2012). This is not the case for other sample mediums, such as faeces, where samples must be frozen quickly and kept at -20°C to avoid bacterial or microbial degradation of the FGM (Sheriff et al., 2011).

As with all measures of GC, some consideration of the effect of confounds is necessary when assessing cortisol from the hair. Cortisol deposition in the hair may be altered by the body location from which the sample originates and even by the colour or type of hair (guard or undercoat) used (Macbeth et al., 2010; Otovic & Hutchinson, 2015). Care should be taken with sampling methods to avoid contamination from other biological mediums and from hair outside of the sampling area, where appropriate to the methods.

1.4.3 Heart rate variability

Upon encountering stress, the autonomic nervous system (ANS) is activated in order to restore homeostasis (Moberg, 2000). The ANS regulates many biological processes (e.g. heart rate). This is achieved by the actions of its two counteractive components, the sympathetic nervous system (SNS), responsible for the activation of the 'fight or flight' response, and the parasympathetic nervous system (PNS), often referred to as the 'rest and digest' system (Wehrwein et al., 2016). During homeostasis, the ANS is focusing on the regulation of internal viscera. However, in response to stress, the ANS shifts its regulatory response to prioritise external stimuli (Kim et al., 2018). This results in a withdrawn PNS response, which means less inhibition of the SNS and thus, an increased SNS response. This will affect alterations in key processes in response to stress, such as changes in the cardiovascular and respiratory system. This leads to, for example, increased heart and breathing rates which are often associated with stressful situations (Broom & Johnson, 1993; Moberg, 2000).

Heartbeat activity is primarily regulated by the sinoatrial node (SN), which is under the control of both the PNS (sometimes referred to as vagal) and SNS. At rest, cardiac activity is controlled by the reciprocal actions of the SNS and PNS, but vagal regulation dominates (von Borell et al., 2007). The PNS effects to slow the heartbeat and the SNS works to increase it. The rise in the heart rate (HR) typically seen in response to stress could be driven by an increase in sympathetic activity, a decrease in vagal activity or concurrent changes in both. Therefore, measures of HR alone can only provide information on the net effects occurring in the ANS and cannot accurately assess the sympathovagal balance. Heart rate variability (HRV) is a more reliable measure of the autonomic regulation of cardiac activity. Assessment of HRV reflects this sympathetic-parasympathetic balance and shows the heart's adaptability in changing circumstances, highlighting its ability to maintain homeostasis. Typically, a variable heart rate is normal and expected from healthy cardiac function (von Borell et al., 2007). Greater variability offers the flexibility for adaptability in unpredictable and changing circumstances. Low HRV represents poor homeostatic control of the ANS, reducing the ability to cope with stressors (Kim et al., 2018).

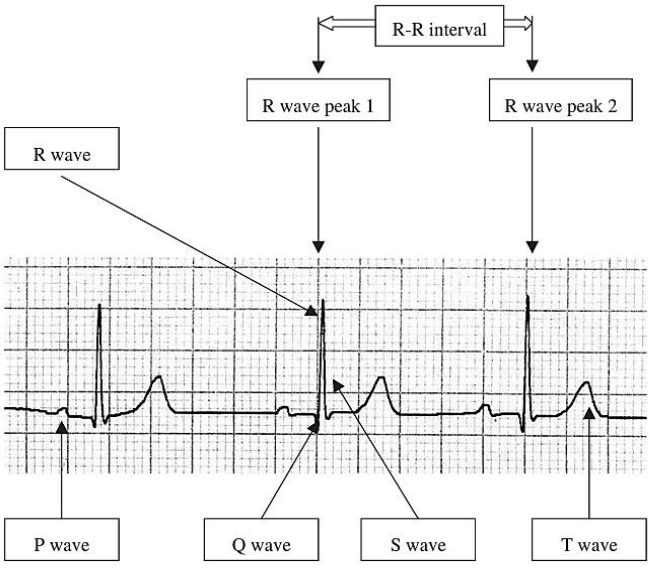

Figure 4: Heart rate as seen on an ECG (Reed et al., 2005)

HRV represents the variation in time intervals between consecutive heartbeats (Karim et al., 2011). It is typically measured using an electrocardiogram (ECG), which records the electrical activity of the heart via electrodes that are attached to the surface of the skin (InformedHealth.org, 2019). During ventricular depolarisation, the QRS complex is formed and at the peak of these are the R-waves. Inter-beat intervals (IBI) are calculated from the distance between consecutive R-waves and these represent the HRV (see Figure 4). IBI may also be referred to as RR intervals or NN intervals (representing only normal R peaks, where abnormal R waves are removed).

Exposure to emotional and physical stress is accompanied by a reduction in HRV (Karim et al., 2011). This can be as a result of activation in the SNS, a decline in vagal response or a combination of the two. Selecting the appropriate variables to analysis can help to determine the influencing factors in this response. There are a vast number of HRV variables that can be used to quantify this balance. Most commonly, time-domain and frequency-domain indices are used. Time-domain measurements assess variations in the NN intervals over time (Shaffer & Ginsberg, 2017). The standard deviation of the NN interval (SDNN) is an example of a time domain measure. SDNN increases with a large and irregular HRV, making it a valuable measure of stress resilience (Kim et al., 2018). Frequency domain measurements assess the distribution of variance as a function of frequency and allows for distinction between the activity of the SNS and PNS. For example, high-frequency (HF) bands reflect vagal activity and low-frequency (LF) reflects SNS activity (Kim et al., 2018).

Due to its dominance in cardiac control during rest, a decrease in vagal activity is a key measurable parameter used in stress assessment. In addition to this, stress resilience can be observed by determining basal autonomic states. Therefore, vagal tone has been used to determine an individual's ability to respond to stress. Individuals displaying stress vulnerability

have been linked to lower vagal tone and those with better adaptability linked to a higher vagal tone (von Borell et al., 2007).

Measures of HRV have been used as a means to evaluate stress in humans and many managed animal species, particularly livestock. For example, HRV changes in horses suggest stress during road transport which increased with journey length (Schmidt et al., 2010) . Changes in HRV parameters indicative of chronic stress were observed in response to clinical diseases, such as lameness in dairy cows (Kovacs et al., 2015) and footrot in sheep (Stubsjoen et al., 2015). HRV can also be used to assess emotional well-being. Variations in HRV have been associated with different emotional states (Catipovic-Veselica et al., 1999) and higher HRV has been observed alongside better emotional regulation (Mather & Thayer, 2018).

1.5 Aims of the thesis

The primary aim of this project is to develop and validate new methods for measuring key physiological and behavioural parameters and discuss how these can be applied, to assess the impact of scientific procedures and the life-time well-being of rhesus macaques (*Macaca mulatta*) used in neuroscience research.

As a study into animal welfare, a key component of the methods was that they were as minimally invasive and unobtrusive as possible in order to avoid further welfare impact. This had the additional benefit of avoiding disruption to the macaques' scientific protocols, allowing data to be collected alongside existing neuroscience practices. This type of 'piggy-backing' research is common amongst welfare research in primates, as there is a reluctance to use any additional NHPs solely for the purpose of welfare assessment (Lemon, 2018). In addition to being minimally invasive, the welfare assessment protocols used aimed to provide a quantitative, objective, and continuous measure of welfare.

Physical activity and behavioural repertoire are examined through the application of accelerometers, an automated and reliable method to obtain continuous data about the activity budget of individual animals. The collection of multiple biological samples is conducted in order to establish the optimum method for detecting the stress hormone, cortisol, within the laboratory environment. Furthermore, this project developed and validated a novel, non-invasive ECG device that can be used to collect heart rate variability (HRV) data, a well-known physiological measure of stress and welfare in multiple animal species.

This thesis will outline the development of these welfare parameters and highlight their usefulness as tools that could be implemented by NHP researchers, providing better insight into the welfare of their animals, and helping to honour their commitment to the 3Rs. It will also discuss the welfare impact of a common neuroscientific procedure, the administration of general anaesthesia, and the cumulative experience of laboratory housed NHPs.

Chapter 2 - General methods

All animals contributing to these studies were enrolled in experimental neuroscience studies under appropriate project licence (PPL) authority. PPLs were issued by the Secretary of State in accordance with the Animals (Scientific Procedures) Act 1986, A(SP)A. All PPLs underwent ethical review by the local Animal Welfare Ethical Review Board (AWERB) and Animals in Science Committee (ASC).

This thesis was undertaken as part of a program of work studying the welfare of macaques enrolled in experimental neuroscience research. To that end, it was essential that data collection for this study did not alter the experience of the animals in their neuroscience experiments. This was important for two reasons. If sample collection would be crossing the A(SP)A threshold ('lower threshold') to be categorised as a regulated procedure, this would have created a potential confound for collecting robust and reproducible welfare data. A procedure is regulated if it is carried out on a protected animal and may cause a level of pain, suffering, distress or lasting harm (PSDLH) equivalent to, or higher than, that caused by inserting a hypodermic needle according to good veterinary practice. Secondly, the primary purpose of neuroscience must be achieved to ensure compliance with the relevant PPL(s). Development of suitable methodology as listed below, was carefully established to ensure that both objectives were met at all times.

All data from NHPs pertaining to this body of work was obtained opportunistically, 'piggybacking' onto the neuroscience protocols. Working in collaboration with PPL holders, PIL holders, and animal care staff, data were collected around various procedural or husbandry events conducted as part of the existing neuroscience research.

2.1 Accelerometers and collars

To assess macaque activity and behaviour, uniaxial, lightweight, piezoelectric accelerometers (CamNtech Ltd Actiwatch®-mini – diameter: 24mm, height: 7.7mm, weight: 7.5g,

memory: 128kB) were used. Actiwatches were set up using Actiwatch Sleep Analysis 7 software and an Actiwatch-mini reader (CamNtech Ltd) to record activity in either 10 or 30 second epochs (allowing continuous data collection for 15 or 45 days respectively). Actiwatches produce activity counts, which is a generic term that reflects amplitude of the voltage generated by the actiwatch in response to the intensity and amount of movement in all directions. The actiwatch operates at a frequency of 32Hz, meaning that the amplitude is measured 32 times per second and the peak intensity is taken as the count. An activity epoch represents the sum of activity counts within that sampling period (e.g. 10 seconds = sum of 10 x 1 second counts). Actiwatches are attached to a customised soft, Microfiber flat neck collar (Dogline). Collars weighed 40g prior to customisation, during which elements were removed, a pocket for the actiwatch was added and material cut down to create optimum sizing for individual animals. Thus, the weight of fitted collars containing actiwatches varied but were always less than 47.5g. This falls well below the advised threshold at which body movement may be compromised (5% of body weight, (Watanabe et al., 2005)), equating to less than 2% of the body weight of the smallest macaque fitted with a collar.

Collar design and placement were carefully considered to avoid discomfort to macaques and minimal disruption to their primary research protocols. Collars were fitted around the neck of the animal whilst either under GA or under restraint for the purpose of a neuroscientific procedure. For example, when restrained in a primate chair for the purpose of behavioural testing or training. Collars were fitted and removed by appropriately trained personnel. Typically, collars were fitted until the actiwatch reached maximum storage capacity (15 to 45 days depending on which epoch was selected). If further data were required after this period, collars were removed and refitted after replacing with a reset actiwatch, usually on the same or next day. Individual macaques were fitted with collars repeatedly throughout the length of this project (2015-2020) to collect data around different experimental and husbandry related events. Collar size was rechecked at regular intervals to account for growth or changes in body weight and ensure optimal fit.

Fitting of a collar has the potential to cause brief and transient discomfort or distress, below the "lower threshold" of a regulated procedure under A(SP)A. Therefore, the welfare of macaques fitted with collars was carefully monitored via behavioural observation for several days after fitting. Animals fitted with collars for the first time took several hours to habituate to wearing it, often pulling, turning, or scratching at the collar following fitting. Collar manipulation had always ceased by the morning following fitting at the latest. For repeated fittings, this habituation period reduced with each fitting (an example of one macaque's response to subsequent collar fittings can be seen in Figure 5).

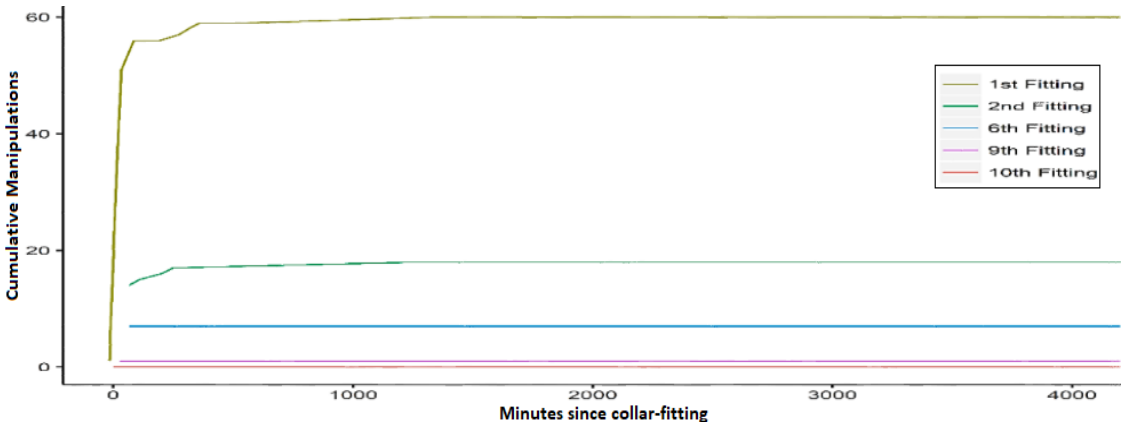

Figure 5: Cumulative manipulations (pulling, turning, scratching) of a collar fitted to a male macaque following subsequent fittings in the same individual.

2.2 Selecting methods of glucocorticoid assessment

To assess glucocorticoids as a measure of stress, the collection of different biological samples were trialled (saliva, urine, faeces, and hair). The aim of this was to establish which methods were best suited to monitor the welfare of rhesus macaques partaking in neuroscience experiments. This included ensuring measures were appropriate to collect within the laboratory environment; could be collected alongside and in collaboration with the neuroscience research, without disruption of their primary goals. Indicators of both acute and chronic stress were desired.

Non-invasive methods of cortisol analysis were trialled to establish how collection protocols fitted into the environment. Non-invasive here reflects the definition for medical procedures 'not involving the introduction of instruments into the body' and not requiring further restraint beyond that used as part of neuroscience procedures. Blood samples were not collected. This was both to avoid an additional stressor, which would have confounded results and potentially impacted the neuroscience protocols, and to negate the performance of a regulated procedure under A(SP)A (Animals (Scientific Procedures) Act, 1986).

The housing system and the daily neuroscience protocols of these macaques allowed for collection of saliva, urine, faeces, and hair samples with minimal disruption to daily routines. In order to obtain samples using the most refined methods and avoid causing potential PSDLH, some sampling techniques required animal training beforehand. This could be done in the home cage, around daily neuroscience testing, using the positive reinforcement techniques (PRT) that the macaques were already familiar with. For further information about PRT techniques, see section 2.4.2.

Measures of cortisol in saliva, urine, and faeces offered opportunity to assess the acute stress response to standard laboratory husbandry, such as social housing, and to neuroscience protocols, such as testing procedures involving restraint. Collection of these samples longitudinally was achievable due to the length of neuroscience studies (often 5-10 years), meaning animals remained within the facility over long periods. Longitudinal data collection and hair cortisol measures could be used to assess chronic stress and accumulation of repeated stressors. Macaques could be easily identified (familiarity with animals by researchers/care staff and distinct chest tattoos), so individual samples could be obtained. Furthermore, sample collection protocols were trialled for their simplicity and ease of implementation.

2.2.1 Saliva

2.2.1.1 Methods

PRT was used to teach macaques the required methods for saliva collection. SalivaBio Children's Swabs (Salimetrics) were used and these are a synthetic material designed to improve volume collection and increase participant compliance, and validated for use with salivary cortisol. Some training was done using flavoured swabs (either tropical/sugar or coconut) and some done using plain swabs.

Tropical/sugar solution was made in line with the SPIT method previously used for cortisol analysis in infant macaques (Roma, 2006). It was 200g of sugar, 500ml of water, and 1ml of LorANN oils tropical punch flavouring. Coconut swabs were created using Vitacoco coconut water. Swabs were soaked in flavouring and then left at room temperature to dry for 2-3 days before being stored in the refrigerator (4°C) until use.

No food was given at least 1 hour before collection, including edible treats and forage. Water bottles were removed from the home cage 30 minutes prior to collection. This was done to avoid compromising the sample, as suggested in the saliva collection protocol from Salimetrics (Salimetrics, 2023). Macaques chewed on or held the swab in their mouths for 1-2 minutes to obtain the optimal saliva volume required for analysis (75µl). Following saliva collection, swabs were checked for traces of blood and discarded if any visible evidence was found.

Swabs were then placed into the swab storage tube (included with SalivaBio swabs), labelled and placed into a freezer at -20°C for storage until shipment for assay. Saliva was assayed at Anglia Ruskin University Psychology Laboratory (previously Salimetrics UK). When thawed, samples were spun at 3000RCF for 15 minutes before being analysed using Salimetrics Cortisol ELISA kit (full method can be found in the manual - (Salimetrics, 2021)).

2.2.1.2 Findings and conclusions

From April 2015 to February 2017, saliva samples were collected from macaques when opportunity allowed. Some attempts at sampling were unsuccessful due to having to discard the swab due to visible blood contamination (no signs of pain from the individuals) or the macaques removing the swab from the handler, resulting in contamination or breakage (individual came to no harm).

Of the samples successfully obtained, 139 samples were part of a trial round, collected from 23 macaques (22 males, 1 female). In this instance, animals were not all fully trained for saliva collection but rather given the opportunity to chew a tropical/sugar flavoured swab freely. The aim of this was to see how viable saliva collection methods could be. Several issues arose from this trial, which allowed for adjustments to be made and saliva collection protocols to be refined. For example, 31% of those samples had insufficient saliva volume to be assayed and 12% had an unusually high cortisol concentration, suggesting they had been contamination with blood or urine. From this, it was clear that animals required training to ensure sufficient saliva volume and to monitor contamination risk. This trial study also highlighted issues with the use of tropical/sugar flavoured swabs. The swabs themselves were highly palatable and desirable to the macaques. This meant that when attempting to collect the samples using these, often the subjects would steal the swabs from the researcher's hand, making them difficult to retrieve and at risk of contamination. Additionally, upon assaying samples collected using these swabs, the lab had issues with being able to pipette the samples. They reported that the samples were highly viscous, so their automated system was not able to pipette the sample but it was possible to assay them using a manual method. This method was more time consuming and more expensive. It was due to these issues that swabs flavoured with coconut water were used, as an alternative approach. This method was recommended in a personal conversation with the Gunnar lab (University of Minnesota, May 2016), who were using coconut water in some of their research and reporting it was working well, with no effect on cortisol concentration. Macaques still found

the coconut swabs palatable but were less likely to take them, overall being more cooperative with the collection method. Also, there were no issues with viscosity reported with the coconut swabs.

Further issues were encountered with saliva sampling as we continued to trial sample collection protocols. Discolouration of some saliva swabs was noted and was suspected to be caused by the food dye that was being consumed by some macaques for the purpose of faecal sample identification (see section 2.2.3.1). This discolouration was observed even though no food was eaten immediately prior to sample collection. As food dye in saliva samples hadn't been encountered before, this led to some confusion about potential confounds when assaying. Furthermore, blood from the gums was frequently contaminating the swab, making them unusable.

A pair of macaques, housed together, were engaged in collection protocols in the early stages of development. Their data can be observed in Figure 6. The protocol for sample collection for this pair involved a researcher entering the room around 30 minutes prior to sample collection to remove ad libitum water. Samples were collected in the morning, between 07:45 and 11am. Macaques remained socially housed whilst obtaining the sample, which resulted in the dominant of the pair, WO, engaging in sample collection for the first 10-15 minutes and then animal WI engaging afterwards. A general observation is that salivary cortisol in WI is higher than that seen in WO. Although this could be attributed to several factors (e.g. social rank), there was a concern that we were just observing an anticipation effect for engagement with the sample collection. Cortisol peaks in saliva are detected approximately 20-30 minutes after a stressor (Sheriff et al., 2011). As WI was always sampled second, and sampling took approximately 10-15 minutes per animal, it could be that WI's cortisol reflected arousal to engage. This is a potential confound that limited the use of salivary cortisol in this study, as resources didn't allow for animals to be sampled at the same time.

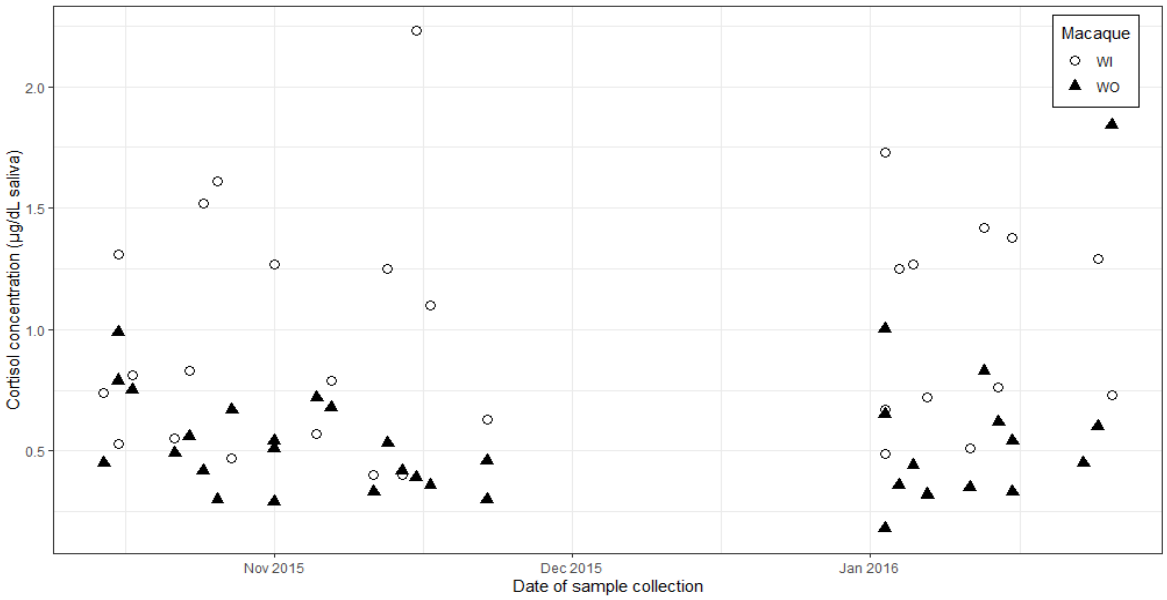

Figure 6: Example of salivary cortisol from two pair-housed male macaques (WI and WO).

There are additional confounds to be considered that were avoided for the purposes of this study. Particular food and drink items may effect salivary cortisol, as is seen with acidic foods for example (Salimetrics, 2023). As PRT plays a key role in macaque's daily testing and training regimes, the ability to deny them food or drink is not straightforward and likely off-putting to researchers. The strong diurnal rhythm observed in salivary cortisol (Cross & Rogers, 2004) would make it difficult to assess certain stressors. For example, if wanting to measure the effect of restraint on salivary cortisol pooling results from macaques may not be possible as individual testing occurs at different times of the day.

2.2.2 Urine

2.2.2.1 Methods

PRT was used to teach macaques the required methods for urine collection. Macaques were taught to come to the front of their home cage and allow a urine sample pot to be held in place in front of their genitals until they urinated. If a sample was not provided within a few minutes, collection protocol would cease and be repeated in 10-20 minutes time. Training of the

collection protocol was relatively quick and straightforward, with all individuals at least stationing and allowing pot placement after 1-2 weeks of 10 minute daily sessions. However, teaching the final stage of urination was more difficult and time-consuming, requiring waiting around and keeping the animal in an accessible position to capture the behaviour spontaneously. The behaviour would need to be repeated several times in order to be reinforced.

In total, 206 urine samples from 6 male macaques were collected from June 2015 until February 2017. After 2-3 weeks of daily training, 2 of these 6 macaques would reliably provide a sample every time and would obviously be trying to urinate when seeing the collection pot. The other 4 macaques learned to station for sampling during this time but collection of samples was reliant on opportunistic urination. Samples were labelled and placed in a freezer at -20°C. Once animals were trained, urine samples were quick and simple to collect. As they could be collected from the home cage, there was no disruption to their primary research goals.

2.2.2.2 Findings and conclusions

Due to time and resource constraints, urine samples were not assayed and sample collection protocol ceased. Urine sampling could be an effective tool for monitoring HPA axis activity. However, the time required for animal training is significant and therefore, the urine collection may need to be the primary focus of the research project or have sufficient resources dedicated to it in advance. As only 2 of the 6 macaques reliably produced a sample, employing this method long-term would require either more investment in training or an alternative method to successfully obtain samples each time. The macaques used in this study were already partially trained before urine protocol training, i.e. already familiar to PRT, were accustomed to researcher presence and used to coming forwards in the home cage. Attempting to teach this collection protocol to less experienced animals may be more difficult. Furthermore, although the animals were well-trained to station in one place, they were still 'freely moving,' i.e. not restrained. This means that researchers responsible for collecting urine samples need to be well trained and able

to understand the animal's behaviours and movements, as a safety precaution. The freely moving element also means that animals could move during urination and so a urine sample may not be collected in its entirety. This may have some effect on the concentration of cortisol within the sample (Sheriff et al., 2011).

2.2.3 Faeces

2.2.3.1 Methods

Samples were collected opportunistically, aiming to capture key events and experiences within the career of a neuroscience macaque. Where possible, individual faecal samples were collected daily up to 2 weeks before and after significant events. Events included, but were not limited to, experimental procedures, such as administration of a general anaesthetic, and husbandry related events, such as change in social housing (gain or loss of social partners) or relocation to a new holding room. As with other data collected during this welfare project, the 'piggy-backing' nature of this research did sometimes prevent data collection from occurring at the optimal time.

Group-housed animals were given 1-2 tsps. of Rainbow Dust ProGel® food colouring (blue, green, or grey) in their daily primate mash to colour faecal samples, for individual sample identification. Food colouring is a commonly used indigestible faecal marker and was selected as it was a safe, non-toxic ingredient that had no or minimal taste (Fuller et al., 2011). The addition of dye into food did not seem to effect the macaque's experience of their food. Food intake was unaffected by the addition of dye and observations, including feedback from animal care staff, revealed no differences in the macaque's engagement with food when it contained dye. Faecal samples were collected in the morning (between 07:00-12:00) by animal technicians, with protocols given to collect the freshest sample available whilst avoiding samples with obvious contamination (for example, from urine). Whole samples were placed inside plastic storage containers suitable for biological materials (Rotilabo PP 20ml) and labelled. Following collection,

samples were placed into the freezer (< -20°C) until shipment for analysis. In total, 9434 faecal samples were collected from 64 macaques (61 males, 3 females) from January 2015 to March 2020.

In total, 1716 samples (n=36, all male) were sent for cortisol assay. Samples were shipped in batches on dry ice in less than 24 hours from Oxford to the Deutsches Primatenzentrum (DPZ) in Gottingen, Germany. Faecal samples were then assayed for cortisol metabolite using an enzyme immunoassay (EIA) technique. EIA uses antibodies specific to a desired molecule, in this case a cortisol metabolite (11ß-hydroxyetiocholanolone), to determine the amount of that molecule present in the sample. The antibodies are bound to the base of each well within a microtiter plate. Prior to adding to the well, samples are mixed with a labelled version (enzyme-conjugated) of the desired metabolite. When this mixture is added to the wells of the plate, the metabolite within the sample competes with the labelled metabolite to bind to the antibodies. The two metabolites will bind, on average, in the same proportion as they are present within the sample mixture. Following this, a substrate is added to the wells which will produce a measurable colour change in the solution which will be proportional to the amount of labelled metabolite bound to the well. This colour change is therefore inversely proportional to the amount of metabolite present in the sample and the solution colour can then be compared to standards (run alongside samples) to determine metabolite concentration within a sample. The specific protocol for detection of FGM in this study are as follows (assay description provided by Michael Heistermann at the DPZ):

Faecal samples were lyophilized and pulverized and an aliquot (50–70 mg) of the faecal powder was extracted with 3 ml 80 % watery methanol by vortex-mixed for 15 min (Heistermann et al., 1995). All faecal extracts were analysed for cortisol metabolites using a group-specific microtiter plate EIA for the measurement of immunoreactive 11ß-hydroxyetiocholanolone, a major metabolite of cortisol in the faeces of primates (Heistermann et al., 2006), including the

rhesus macaque (Higham et al., 2013).The EIA used an antibody raised in a sheep against 5ß-androstane-3alpha,11ß-diol-17-CMO-BSA and 5ß-androstane-3alpha,11ß-diol-17-CMO-DADOO-biotin as label (Ganswindt et al., 2003). The assay has been successfully applied to monitor adrenocortical activity in a variety of primate species of all major taxa (e.g. (Heistermann et al., 2006); (Fichtel et al., 2007)), including several species of macaques ((Ostner et al., 2008); (Girard-Buttoz et al., 2009); (Young et al., 2014)). It has previously also been validated for assessing adrenocortical activity in rhesus macaques in response to trapping stress (Hoffman et al., 2011) and has been applied successfully to examine stress reactivity in response to social stress free-ranging animals (Higham et al., 2013).

Hormone determinations were carried out as described in (Heistermann et al., 2004). In brief, faecal extracts were diluted 1:10 to 1:1000 (depending on concentration) in assay buffer (PBS, pH 7.2) and 50µl aliquots taken in duplicate to assay. Samples and 11ß-hydroxyetiocholanolone standard (50 µl; dose range: 0.6-156pg) were combined with the labelled hormone and antiserum (50 µl each), and the plate incubated overnight at 4°C. After incubation, the plates were washed four times and 150µl of streptavidin-peroxidase was added to each well and the plates incubated at room temperature for 60 min in the dark. Following a second washing step, 100µl of TMB substrate solution (Fisher Scientific, Art. No. 34029) was added to each well and the plates incubated on a shaker for another 60 min at room temperature in the dark. Absorbance was then measured at 450 nm (reference 630 nm) in a spectrophotometer. Sensitivity of the assay at 90% binding was 1.0 pg. Intra- and interassay coefficients of variation of high and low value quality controls were 5.7% (high) and 7.9% (low) and 10.6% (high), and 10.8% (low), respectively. All hormone concentrations are expressed as ng/g faecal dry weight.

2.2.3.2 Findings and conclusions

The protocols required for faecal sample collected were simple and easy enough, such that the collaboration between the welfare research team, the animal technicians and the

neuroscience researchers allowed for successful collection of thousands of samples. Each collaborator played a key role by performing a quick task that was minimally disruptive to their daily routines. Within these protocols, the welfare team provided logistics and organisation, e.g. instructing which samples to be taken and when, providing the pre-labelled pots and cataloguing samples post-collection. The animal technicians collected samples alongside routine husbandry and facility maintenance, following guidance on proper collection protocol. The neuroscience researchers kept the welfare team updated with information about experimental procedures and were able to facilitate sample collection by providing the dye for sample identification. Their contribution to this work was essential to the success and the collaborative experience was a positive one.

Whilst sample collection may be easily implemented, the ability to store samples, as well as funding and resources required for analysis of the samples, do offer some limitations to the incorporation of the measures. Faecal cortisol is at risk of degradation, so it is essential to store and maintain samples at less than -20°C until assaying. This requires the resources to store samples at this temperature and, if samples are shipped for assaying, maintenance of these temperatures during shipment. Unfortunately, we did have a batch of faecal samples shipped that could not be assayed, as they had begun to defrost before arrival at the lab. Therefore, it would be recommended to separate out samples, so a data pool is not lost if an incidence like this should occur. However, this would come with additional resource costs to split and store samples in this way.

2.2.4 Hair

2.2.4.1 Methods

Hair samples are obtained by shaving an area, using standard animal clippers, at the nape of the neck, approximately 5cmx5cm, whilst animals are restrained or under GA as part of their neuroscientific study. The neck was chosen as it was an accessible area during neuroscientific

restraint in a primate chair, from which the animals head protrudes and is fixated in position (see section 2.4.3 for further details). Also, this area is less accessible to the macaque, reducing the risk of contamination from self-grooming (Meyer & Novak, 2012). It is possible that animals could have been trained for hair sample collection but this technique is likely to be less reliable. Without restraint, the macaque could move during sampling, making the collection less precise and posing a risk to the collector. Samples were collected using a shave-reshave method. This means that the same area is shaved every time with only hair taken from the same sample area kept for analysis. This ensures a known time period during which cortisol in the sample will have accumulated (Meyer & Novak, 2012). Shaving was conducted by an appropriately trained individual. When shaving, clean gloves were always worn and clippers were thoroughly cleaned (blade removed, brushed and sprayed with clipper cleaner) after each animal. Samples are stored in aluminium foil at room temperature until shipment for assaying. In total, 868 samples were collected from 45 macaques (4 female, 41 male) from February 2016 to March 2020. As the methodology for hair sampling required the animals to be restrained, collection was restricted to macaques that had been fitted with a head fixation device (as part of their neuroscience research – see section 2.4). Macaques could not start hair sample collection until after the implantation of the headpost (plus a recovery/habituation period). Therefore, the time engaged in sample collection varies between animals, with the longest being 4 years and the shortest is 3 months. Every 6-8 weeks, hair samples from all animals on study were assessed for regrowth and collected within 1 week, if required. For each macaque, samples were collected until they could no longer participate (e.g. due to termination or removal of the headpost) or the study ended.

A pilot study to check that cortisol could be detected in the samples was conducted in 2017. In this instance, 78 samples (8 males) were sent to Jerrold Meyer's lab at the University of Massachusetts Amherst where the samples were assayed using an EIA kit (see reference for methodology, section 'Method 1 - Large Samples' - (Meyer et al., 2014)). Due to time and budgetary restraints, no more samples could be assayed at that time.

In January 2021, a collaboration was established with Dr Esther Carlitz and Professor Clemens Kirshbaum (TU Dresden), both experienced with hair cortisol research, including studies using primates. Thus, 570 samples (n=28, 24 male and 4 female) were sent to the Faculty of Psychology at TU Dresden for LC-MS/MS assay (see reference for methodology (Gao et al., 2013)). All samples were sent and assayed at the same time. Samples from females were sent as a reference and to determine detection of other hormones, so results from those 4 were not used in cortisol analysis.

2.2.4.2 Findings and conclusions

Hair sample collection methods provided adequate quantities of hair for analysis. There were no observed issues with sample contamination, as the shaved area was always large enough such that there was confidence that the collected sample only contained hairs that had been previously shaved. Furthermore, samples were immediately placed in foil for storage once collected and hair appeared clean when visually assessed prior to collection.

Hair sample collection protocols were easily applied with collaboration from the neuroscience researchers. However, attempting to collect hair samples from all macaques within 1 week, working around their individual testing schedules, was logistically challenging. Whilst the shaving was quick, macaques were tested in different rooms and at different times throughout the day. As sample collection had to be completed by a trained individual, this usually meant lots of close collaboration, time management and adaptability, particularly as macaque testing times could be unpredictable. Whilst this was manageable, a future method whereby neuroscience researchers are appropriately trained to collect hair samples from their own animals during testing sessions may be more appropriate.

2.2.5 Summary

All sample types were valuable as potential indicators of acute and chronic stress. Whilst all methods were non-invasive and minimally disruptive, some methods required more time and resources than others.

Urine and saliva sampling required investment into animal training and obtaining samples successfully was less reliable than faecal or hair sampling. Furthermore, saliva sampling was most frequently subjected to confounds such as contamination and anticipation effect. As the nature of this study was to obtain welfare measures alongside neuroscience research, in collaboration with researchers and animal care staff, the simplest methods were most manageable to implement. Therefore, only faecal and hair sampling methods were adopted for acute and chronic stress assessment and some findings from these measures can be found in 'Chapter 5 - Cortisol and non-human primate welfare.'

2.3 Developing a neck-based ECG device

The concept of the neck-based ECG was developed through discussion between Dr Caroline Bergmann (University of Oxford) and Keith Simpson (previously Vetronic Services Ltd, now Burtons Medical). The project was initiated through funding from the Wellcome Trust Strategic Award. The ECG was designed through this collaboration, with Dr Simpson conducting manufacturer and technical development. As this is a novel design, the main focus of this project has been to develop a proof of concept. Development was an iterative process, involving the ECG being sent back and forth between Oxford and Vetronic Services for further adjustments following feedback and data collection from practical trials.

Further information about the design and development of this ECG device can be found in 'Chapter 6 -Development of a novel ECG device for measuring heart rate variability.'

2.4 Neuroscience methods

The following section describes the housing and husbandry conditions and the neuroscience protocols that the macaques used in this study were enrolled in. This gives a general overview of the experience for all animals. Most practices are the same but there may be some differences in training and testing regimes, as macaques were engaged in different study protocols (under different PPLs) with differing scientific aims. All procedures were conducted under licenses from the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986.

2.4.1 Animal housing and husbandry

All housing and husbandry conditions adhere to all sections of the 'Code of Practice: Animals' for non-human primates (NHP) (Home Office, 2014a). In this section, I am embracing the need for transparency and thorough reporting of research involving animals by complying with the Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines (Percie du Sert et al., 2020).

2.4.1.1 Room and cage design

Rooms are 24.65m^2 and will hold a maximum of 10 NHPs. Cages are 2.2m high (floor to ceiling) and range from 2.2m^2 to 3.65m^2. Cages are assigned at a minimum of 1 cage per animal. Cages are situated in a horse-shoe formation around the edges of the room, giving a central corridor for human access and direct visibility of all enclosures. Cages are separated by moveable, but lockable, plastic or wooden boards to permit the movement of animals around the cages. There are two enclosure types, which are referred to as 'rectangular cages' and 'pens'. Rectangular cages have a crush-back mechanism to allow restraint but have limited physical enrichment due to this design (avoiding obstruction of the crush-back). Pens are larger and include a range of enrichment such as swings, fire hose, and ladders. All macaques will have

access to at least one pen but for some parts of the day (e.g. when social groups are separated for neuroscience testing), some macaques may be restricted to a rectangular cage. All rooms were served by a central corridor and had windows that looked out into this corridor.

2.4.1.2 Cage cleaning

At a minimum, cages are cleaned once per week. Either a 'dry clean,' which involves removing soiled substrate, surfaces wiped, and enrichment cleaned. Alternatively, a 'wet clean' which involves emptying the room of all substrate and enrichment and pressure washing. Wet cleaning occurs at a minimum of fortnightly.

2.4.1.3 Social and group-housing

NHPs are group-housed in pairs or small groups, having direct, tactile access to conspecifics. Single housing is only permitted under exceptional circumstances, on health or welfare grounds. Multiple NHP groups will be housed in the same room and so all NHPs, single or group-housed, have socialisation opportunities available in the form of visual, olfactory, and auditory access to conspecifics. Windows allowed for further visual access to NHPs housed in neighbouring rooms.

2.4.1.4 Environmental conditions

Lighting: Facility lighting operates on a 12-hour light/dark cycle. Light hours are from 7am to 7pm and include a simulated dawn and dusk period. A skylight down the central corridor between rooms permits some natural light into the facility.

Temperature: Holding rooms are maintained between 15-24°C.

Humidity: 40-70% humidity is maintained were possible, with some fluctuations occurring during wet cleaning of the rooms.

Ventilation: Mechanical ventilation operates at a rate of 10 to 12 air changes per hour.

2.4.1.5 Food and water protocols

Fluid control and modification of the diet are permitted under the authority of the PPL. This is to allow food and fluid rewards to be given as part of the neuroscience testing and to ensure animals are sufficiently motivated to work for such rewards. As a standard, animals are fed a specified volume of 'primate mash' each day (specific to individual animals) and one portion of fruit or vegetable (agreed by NACWO). 'Primate mash' is prepared daily and consists of powered primate diet mixed with water, omega oil, raisins, and forage seeds. The portion is split on non-testing days to provide additional enrichment. On testing days, it's one portion given as a jackpot reward following the testing session.

Unless required for scientific protocols and specified in the PPL, water was provided ad libitum. Fluid control protocols were often permitted by the PPL but were used sparingly. Alternative approaches were attempted to motivate behavioural performance prior to implementing restrictions, such as the provision of a more desirable reward (e.g. smoothies or diluted juice instead of just water). There were instances (e.g. when being asked to perform a particular long or complex task or when behavioural performance dropped) when animals did require fluid control protocols in order to achieve the scientific objective. When it was required, water access was controlled such that no access was permitted in the period prior to behavioural testing and ad libitum access was given in the period immediately after testing. Minimum requirements of ad libitum access varied with PPLs but were usually at least 3 hours within a 24-hour period. In instances where water control was required for a macaque but not their social partners, social partners were provided with a minimum of 6 hours ad libitum access. In most cases, these minimum requirements were exceeded, and macaques typically had longer than 3 hours ad libitum water access per day. For example, a macaque tested daily at 9am would typically have fluid reward during their testing session and then ad libitum access upon their return from testing (e.g. 11am) until the end of the working day (4-7pm), at which time access would be restricted again ready for the following testing day. Fluid control protocols were

typically applied only on weekdays (or 'testing' days) and 2 days of ad libitum water was given per week.

2.4.1.6 Environmental enrichment

All cages contain multi-level perches and ropes to facilitate climbing. Substrate (shavings) is provided on solid floors to encourage foraging behaviours. Scattered into the shavings, all animals receive forage (mixture of seeds, grains, rice, peas – small, low calorie items) at least once daily, usually twice daily. Cognitive enrichment is provided regularly in the form of, for example puzzle feeders, kong toys and cardboard boxes filled with forage and hay. Additional enrichment is provided to singly housed animals and to group-housed animals on non-testing days, such as weekends.

2.4.2 Training methods

To engage in neuroscience studies, macaques are trained using mixed reinforcement training (a combination of positive (PRT) and negative (NRT) reinforcement training) and a combination of behavioural shaping and operant conditioning techniques. Typically, PRT was attempted first and NRT was used where several attempts at using PRT had failed. For PRT, macaques are given a command signal and are rewarded with treats (e.g. raisins, peanuts, diluted juice) for producing the desired behaviour. A 'bridge' was used (e.g. whistle or clicker) to connect the desired behaviour with the receiving of reward. For complex behaviours or tasks, macaques are trained in stages, by successive reinforcement of behaviours closer and closer to that of the target behaviour.

NRT involved presenting macaques with a negative stimulus or reducing the space available to them through the use of crush-back mechanisms. NRT was applied incrementally, ensuring the minimum amount of negative reinforcement was applied to achieve the desired behaviour. NRT was conducted quickly and effectively, with the negative stimulus quickly removed following the

desired behaviour and was usually followed by still rewarding the macaque with treats. All training, both PRT and NRT, was performed by appropriately trained individuals.

A study by Stuart Mason and colleagues was conducted on neuroscience macaques at Oxford and reflects the training techniques applied throughout the facility (Mason et al., 2019).

2.4.3 Transportation and restraint

Cages are designed such that researchers and animal care staff do not directly handle the animals. Removable boards between cages and hatches at the front, to which either primate chairs or transport boxes can be fixed, allows for the movement and removal of macaques for daily neuroscience testing.

Transport boxes are typically used to move macaques about the facility but not used for neuroscience testing, with the exception of some behavioural data collection. These are metal boxes, elevated on legs and wheels, with a caged front allowing animals to see out and be handed rewards.

Primate chairs are specifically designed restraint boxes that enable the NHP to 'sit' in place for sustained periods whilst undertaking neuroscience testing. Primate chairs at this facility were 'closed' (i.e. NHP is enclosed inside an acrylic box) (McMillan et al., 2017). Macaques are trained to enter the box through a door and then later present their head through an opening, which is then closed around the neck to fix it outside of the box ('neck plating'), allowing access for brain recordings. In electrophysiology studies, primates were seated vertically in the chair and in MRI studies, the chair was placed horizontally and the NHP seated in a 'sphinx' position.

During brain recording or MRI imaging, the head needed to be fixed in order to effectively and safely record from the brain. To achieve this, macaques were fixed by attaching the animal's headpost to the primate chair ('head fixing'). Headposts required prior surgical procedures to affix to the skull (see section 2.4.7).

2.4.4 Testing regimes

The following will describe a typical daily testing regime of a macaque used in neuroscience research. There may be some differences between protocols and macaques, but the overall routine is the same. All steps of the testing regime are carried out by an appropriately trained researcher, who will hold a valid PIL if conducting regulated procedures. Macaques have been trained to cooperate with each stage of the regime using PRT. However, NRT can be used if required.

Macaques are separated (if socially housed) and encouraged to move voluntarily between different sections of their enclosure. This is to move the macaques to a cage to which the primate chair can be attached (rectangular cage). Macaques will then enter the primate chair, in which they are transported to a testing room. Before testing, macaques were further restrained through 'neck plating' and 'head fixing.'

Testing sessions can be categorised into either training or recording sessions. Neuroscience experiments require macaques to perform complex, computer-based cognitive tasks which they must learn before data acquisition can begin. In training sessions, no brain recordings take place and the macaque is learning (in steps) how to perform the required task. It can take several months or years of repeated training sessions before brain recordings can begin. In recording sessions, the macaque is performing the task they have learnt whilst recordings of the brain are being taken, either via a functional MRI scan or electrophysiology. Testing times varied but typically lasted ~1 hour and recording sessions typically took longer than training sessions.

When performing cognitive tasks, reward is received for correct trials, usually in the form of diluted juice or a smoothie. Upon completion of testing, a 'jackpot' reward was received, which usually consisted of the macaques' daily meal. Head restraint was removed and the animals typically ate in the primate chair before being returned to the home cage, where they were repaired with their social group (where applicable).

Testing regimes were usually conducted on weekdays, with some exceptions but macaques had a least one day a week with no testing. Testing times between macaques varied but usually remained consistent for each individual.

2.4.5 Clinical procedures

Health screening for each macaque was typically conducted on an annual basis. Wherever possible, veterinary examination were performed when a macaque was under general anaesthesia (GA) when required for their respective neuroscientific research protocols. In instances where no GA protocols were performed within a year of the previous health check, a short GA (10-20 minutes) was used to conduct the clinical examination and obtain biological samples (individual health and colony health status).

Due to the social nature of macaques and attempts to group-house wherever possible, macaques did sometimes get injured as a result of fights with conspecifics. All such injuries would always be assessed, and subsequent treatment advised by the named veterinary surgeon (NVS) or deputy. Where required, the assessment and/or treatment of a wound(s) including surgical closure was undertaken during a GA, typically of short duration (<30mins).

Where possible, medication was administered orally by mixing into palatable food items to be consumed voluntarily (hand-feeding). Alternatively, treatment may have required injection by briefly restraining the animal in the home cage using a crush-back mechanism. Topical treatments would also be administered as required and depending on the animal's cooperation. Where possible, PRT was used to administer treatment or animals were restrained using the crush-back mechanism.

2.4.6 MRI under general anaesthesia

MRI under GA was required to determine patterns of rest, interconnections, and functional interactions of brain circuits specific to brain regions specified by the PPL. All anaesthesia

protocols include two steps (i) sedation in the home cage (injection using the crush-back mechanism) to enable safe transfer to the preparation area and (ii) subsequent induction and maintenance of general anaesthesia (GA) appropriate for the nature and duration of the procedure. For MRI, this would typically involve administration of a combination of sedatives (routes – subcutaneously or intramuscularly), typically including ketamine, benzodiazepines, alpha-2 adrenergic receptor agonist, partial opioid agonists (e.g. buprenorphine). Food was removed for 12 hours (overnight) and water was typically removed for ~2 hours prior to sedation. All animals were intubated and ventilated (intermittent positive pressure ventilation - IPPV) as required. An intravenous catheter was routinely placed (e.g. saphenous vein) and fluids administered. Intra-anaesthesia monitoring included all vital parameters, typically heart rate, oxygen saturation, respiratory parameters (e.g. end-tidal CO_2, respiration rate, airway pressure), non-invasive blood pressure, peripheral and core body temperature. Duration of GA varied from 1-9 hours. GA was maintained using volatile anaesthetics (e.g. isoflurane, sevoflurane). Alpha-2 adrenergic receptor agonists were typically reversed during the recovery period, as required, to facilitate a smooth and swift recovery. Macaques were closely monitored until fully recovered and returned to their home enclosure typically within 1-hour post-recovery. Whilst recovering, macaques remained in direct proximity to their social partners, to enable social, visual, and olfactory contact throughout. Once they had consumed their usual daily food intake, they were returned to their social group (typically by the next morning). GA protocols were similar across all PPLs and neuroscientific protocols and are stated in further detail in the published neuroscience research (e.g. (Mars et al., 2011)).

2.4.7 Surgical procedures

Surgical procedures are typically performed for the placement (or repair) of cranial implants, which includes headposts and recording devices.GA procedures are the same as for MRI procedures (see section 2.4.6). In addition, macaques are administered peri-operative

antimicrobial treatment (typically amoxicillin) which is continued for typically 5 days post-operatively or extended as advised by the NVS. Furthermore, analgesia is given intraoperatively e.g. a non-steroidal anti-inflammatory drugs (NSAID) and partial or full opioid agonists (e.g. fentanyl, alfentanil) depending on the nature of surgery. Post-operatively, additional analgesia is administered, e.g. meloxicam (NSAIDs) and partial or full opioid agonists (e.g. buprenorphine and/or methadone for intracranial implants e.g. microarrays), for as long as required as advised by the NVS. These post-operative treatments were standard across all macaque surgeries discussed in this thesis. Further treatment may have been administered under the direction of the NVS, as and when required, to promote recovery and minimise suffering (e.g. additional analgesia such as paracetamol, antiemetic treatment such as maropitant and anticonvulsant such as midazolam). These were given on a case-by-case basis and were not the standard.

The surgical implantation of a headpost involves a skin incision and permanent attachment of a device that allows for the head to be fixed in position when the macaque is restrained in the primate chair. This is required for scientific reasons such as immobilising the animal to enable safe placement of intra-cranial electrodes, acquisition of MRI images, and some sensory neuroscience protocols, e.g. to allow eye tracking. Titanium headposts have a footplate consisting of several 'limbs' arranged in the configuration of a K or circle. Titanium headposts are considered more refined as the 'legs' are bent during surgery to match the curvature of the skull and attached with specialised titanium bone screws, allowing osseointegration of biocompatible titanium into the bone tissue. For animals needing to undertake an MRI, titanium headposts are not suitable due to the artefacts that would be caused on the images. In these animals, headposts are made from acrylic and do not have any legs. Instead, biocompatible screws are used to anchor the headpost and then medically graded bone cement (e.g. Palacos) is used to affix the headpost to the skull. Typically, surgeries lasted up to 3 hours. A headpost removal surgery is also discussed in this thesis and is the reverse of the headpost implantation procedure, essentially

removing the device from the skull and surgical closure of the skin. GA and drug protocols were the same for the removal surgeries as they were for the implantation.

Some studies require placement of either a recording 'chamber,' 'microarray' or 'microdrive.' Chambers are of cylindrical shape and allow the lowering of electrodes or microdrives into the brain, to enable electrophysiological recording of brain activity. The electrodes are removed at the end of each recording session. These implants require ongoing management including routine veterinary examination, maintenance, and treatment (e.g. topical lavage, antimicrobial treatment). Microarrays are permanently implanted in cortical or subcortical structure. They require implantation of a pedestal to the skull, through which the wires connected to the microarray can be accessed for recording. The 3rd category are referred to as microdrives. These involve a 2-stage surgery. Following a previous surgery to implant a headpost, the first surgery involves implantation of a chamber-like structure (microdrive base chamber) onto the skull. During the second surgery, a craniotomy is performed and a 'hairbrush' like multichannel microdrive is permanently inserted and affixed to the recording chamber. The position of the electrodes can be manually adjusted and are lowered into the brain incrementally (e.g. after a number of recording sessions) and cannot be retracted. As there are no nociceptors in the brain itself, the lowering of electrodes and microdrives for recordings should not cause the animal pain. Surgeries to affix all types of cranial implant require a skin incision and craniotomy to allow placement of the recording device onto the skull or, in the case of microarrays and microdrives, cortical or subcortical structure. This is followed by surgical closure. Compared to headpost surgeries, these surgeries are more complex and typically carry a higher risk of infection. Surgeries for chambers typically lasted 3-6 hours, microarray and microdrive surgeries would often be longer, lasting up to 12 hours. Animals will receive multimodal analgesia during and post-operatively, as well as a prolonged course of antibiotics as advised by the NVS (as discussed at the start of this section). To minimise risk of intraoperative intracranial oedema, array and microdrive lowering/craniotomy protocols all included the administration of steroids

(e.g. dexamethasone or methylprednisolone) peri- and post-operatively for several days (following NVS advice). Occasionally, follow-up GAs and/or treatments were required for either neuroscience or veterinary purposes (e.g. due to infection) and these are broadly referred to in this thesis as 'procedural interventions.'

Following surgery, macaques were kept separate from their social groups until fully recovered (typically 1-2 days post-op). Macaques recovered in their home cage with restricted access to begin with. Space was opened up when the animal's recovery permitted it was safe to do so (as decided by the NVS and NACWO). Typically, macaques had full access to their normal enclosure and all of its environmental enrichment 1-2 days post-op. Post-operative drugs (as discussed previously) were administered in the home cage using the crush-back mechanism.

2.4.8 Timeline of procedures

This section aims to give an impression of the chronological experience of a macaque engaged in neuroscientific research at the Oxford facility. Timelines vary between study protocols and often between individuals too, as much of the timeline of events is determined by the behaviour and participation of the macaques in their cognitive tasks.

Following arrival at the facility, macaques are typically granted a settling in period of 1-2 weeks, to allow them to habituate to the new environment. After this, researchers begin working with the animals using PRT and incremental training steps (see section 2.4.2), with the ultimate goal of getting each individual to perform a cognitive task whilst restrained to facilitate either a functional MRI (fMRI) scan or electrophysiology recordings to take place. No surgical procedures are conducted until individuals are able to perform cognitive tasks at a satisfactory level, such that data acquisition (brain recordings or relevant behavioural data) can immediately begin following the required surgery and recovery. Depending on the complexity of the task and other requirements of study protocols (such as MRIs done under GA), it typically takes several years to reach this level.

For macaques engaged in fMRI studies, the only required surgical procedure to meet the scientific aim was the fitting of a headpost device for immobilising the animal for effective scanning. Of the macaques engaged in this welfare study, the time from arrival at the facility to the implantation of the headpost ranged from 1 to 4 years.

MRI procedures conducted under GA are not dependent on cognitive task performance and so were conducted at any time, although not typically within the first 6 months of arrival and not within 1 month of a previous MRI GA. Both animals engaged in fMRI and electrophysiology studies would have experienced at least 1 MRI under GA.

Electrophysiology studies required the fitting of a headpost device and a recording device (microarray, microdrive or chamber – see section 2.4.7) to achieve their scientific aims. Headposts were fitted 1-3 years after arrival in Oxford and recording devices were typically fitted 1-3 years after that. During the time period when a headpost was fitted but brain recordings could not be taken (i.e. before the recording device surgery), macaques performed cognitive tasks under restraint to provide valuable behavioural data. Often the headpost and recording devices could be personalised to better fit the skull of each macaque. In those cases, an MRI under GA was conducted a few months prior to surgical implantation in order to map the 3D structure of the skull, on which the device could be customised.

Chapter 3 - The impact of general anaesthesia on non-human primates

3.1 Introduction

General anaesthesia (GA) aims to create an environment in which the subject is no longer susceptible to external stimuli. It achieves this through the induction of multiple states such as amnesia, analgesia, immobility, and unconsciousness. The anaesthetic state is induced by disruption to the central nervous system, through support of inhibitory neurotransmission and suppression of excitatory neurotransmission. Despite clinical use of GA since the 1800s and similar functional disruption seen across all taxa, the exact mechanisms of action of anaesthetics are still not fully understood, although some actions have been identified. At a molecular level, anaesthetic agents have been shown to disrupt the functioning of ion channels (e.g. potassium, calcium, and sodium) as well as affecting alterations in the cellular cytoskeleton and mitochondria (Kelz & Mashour, 2019; Son, 2010). Such actions cause disruption to neuronal activity, particularly in brain regions associated with sleep, wakefulness, and consciousness, resulting in a controlled unconscious state.

The typical life of a non-human primate (NHP) used in neuroscience research involves exposure to a number of non-natural stressors. The administration of GA is a commonly used procedure among NHP neuroscience models. GA will immobilise the animal and reduces awareness of any stressful or potentially painful scientific procedures. Therefore, GA use is essential to conduct many common neuroscientific procedures. The immobilisation ability of GA facilitates the acquisition of brain images through MRI (Hildebrandt et al., 2008). For this procedure macaques are maintained under GA throughout the length of the MRI, which varies depending on the data required but will usually last 3-9 hours. With a few exceptions required to monitor the animal, such as cannulation, the macaque will experience no tissue trauma during

this procedure. Another common use for GA in neuroscience macaques is to allow the performance of a surgical procedure, such as to provide a cranial implant. Cranial implants may serve as a head restraint (e.g. headpost) or brain-recording device (e.g. electrodes). This allows recording of brain activity in awake animals providing insight into how areas of the brain function whilst performing specific tasks. These implants are chronic and attached to the skull requiring a surgical procedure to fit, which entails tissue trauma.

As GA works by interrupting the central nervous system and disrupting many physiological processes, its administration also has a potential welfare impact. Common short-term side effects from GA include confusion, vomiting, and dizziness, among others (Smith et al., 2021). There is also some evidence of long-lasting effects such as changes within the central nervous system leading to cognitive impairment (Vutskits & Xie, 2016). Although animal welfare is monitored in the hours following any GA procedure via cage-side assessment, further investigation is required to determine longer-term welfare impact and highlight potential refinements.

The field of neuroscience commonly requires macaques to be placed in studies for long periods, often between 5-10 years. During this time macaques may experience the administration of GA on multiple occasions and for various procedural purposes. Therefore, in addition to the potential acute welfare impact of GA, it is important to consider whether the repetition of this procedure has a cumulative effect, contributing to their cumulative experience (Pickard, 2013). Moreover, if an intervention, such as a GA procedure, causes a reduction in animal welfare it is important to consider whether this compromise of well-being is acute or longer lasting, particularly if the procedure is to be repeated. An animal whose welfare has returned to pre-intervention levels before the repetition of that intervention would essentially undergo the same experience again. However, if the animal has not recovered its welfare state at the time of a repeat intervention, a cumulative effect could occur, leading to a further decline in welfare. Evidence of a cumulative welfare compromise has been seen in mice, showing reduced welfare in

the immediate post-anaesthetic period following multiple exposures to isoflurane GA compared to a single exposure (Hohlbaum et al., 2017). Contrarily, a study in horses found evidence for an increased quality of recovery following multiple GA exposures suggesting habituation to the procedure that may minimise welfare compromise after each event (Platt et al., 2018).

To measure changes to NHP welfare following exposures to GA, an appropriate welfare parameter needed to be selected. It must be able to be quantitatively collected over several days or weeks and the collection method should be minimally invasive. Measuring changes to physical activity levels fit these requirements. Physical activity is an important contributor to both physical and mental well-being, with changes in human activity levels showing association with pain or stressful experiences (Fox, 1999; Stults-Kolehmainen & Sinha, 2014). Elevated stress has been linked to a decline in physical activity and increased sedentary behaviour. However, high levels of physical activity are associated with lower rates of depression and anxiety. A decline in physical activity levels immediately following a GA would be expected as the anaesthetic agent takes some time to wear off. Additionally, studies in several species, including humans and rodents, have recorded a decline in activity levels several days, or weeks, following surgical intervention (Bourque et al., 2010; Hendrickx et al., 1984; Ratcliffe et al., 2021; van der Meij et al., 2017). This highlights a need for extended monitoring, of several days or weeks, to assess long-term changes to activity. However, the change in activity observed in these studies may be associated with more than the GA as additional aspects of surgery may contribute to the decline, such as post-operative pain or allowing wound healing. Due to the limited situations in which a GA would be administered without surgical stress there is little evidence to show the influence of GA alone on physical activity.

Accelerometers can provide a quantitative estimate of the wearer's physical activity by approximating movement intensity and duration (Brown et al., 2013). In humans, accelerometer recordings have been compared and validated against other methods for monitoring physical

activity, such as self-reporting and measures of heart rate (Patterson et al., 1993). Therefore, accelerometers have proven to be an objective and reliable measure of physical activity allowing automated, continuous sampling over periods of days or weeks. Accelerometers have been employed as tools to measure the long-term recovery of patients following surgical procedures, by monitoring the return of their physical activity levels to their individual pre-surgical levels (Bisgaard et al., 2002; Jonker et al., 2020; Ratcliffe et al., 2021; van der Meij et al., 2017).

Furthermore, accelerometers have been validated and used on numerous different species as a means of assessing animal welfare. For example, these devices can provide gait assessment in dairy cattle to detect lameness and detect physical response to pain caused by antler removal in elk (Chapinal et al., 2011; Thierman et al., 1999). A study in dogs used accelerometers to compare activity following surgeries of differing invasiveness and found significant decrease in post-surgical activity following open but not laparoscopic surgery (Culp et al., 2009). Additionally, several studies have applied accelerometers to measure the activity and behaviour of several NHP species (Fehlmann et al., 2017; Sha et al., 2017; Sullivan & Cameron, 2010).

Several prior studies have used accelerometers to assess animal welfare and NHP behaviour. However, this experiment used specifically designed unique collars to suit data collection from freely-moving, socially-housed rhesus macaques undergoing neuroscience protocols. This design and the applied methods aimed to minimise further impact on animal welfare and avoid disrupting the animal's primary research purpose, contribution to neuroscientific studies.

In this study, accelerometers were used to quantify any change in an individual macaque's physical activity following the administration of a GA for two different scientific procedures: MRI or a surgical procedure to affix a head-post to the skull. Time taken to return to pre-intervention activity levels will be assessed and compared. The null hypothesis for this study are:

1. There will be no change in macaque activity following the administration of GA for the purpose of either an MRI or surgical event.
2. There will be no difference between GA administered for the purpose of MRI (NHP experiences no tissue trauma) and surgery (NHP experiences tissue trauma).
3. The length of the GA for MRI events will have no impact on the time taken to return to baseline activity levels.
4. There will be no cumulative effect of repeated MRI events on the activity levels of macaques, suggesting no effect to the animals' cumulative experience.

3.2 Methods

Activity data from 24 male rhesus macaques (*Macaca mulatta*) were collected for this study. Information of housing and husbandry of the animals can be found in section 2.4.1.

3.2.1 General anaesthetic procedures

Between June 2015 and December 2018 data were collected before and after each macaque underwent at least one GA procedure. GA procedures were part of neuroscientific protocols and were conducted for the purpose of either running an MRI scan or for the surgical attachment of a head-fixation device to the skull of the animal. See 'General methods' sections 2.4.6 and 2.4.7 for further details about the procedures. In total, 48 MRI events (macaques = 19) and 16 headpost surgeries (macaques = 16) were monitored. Exclusion criteria apply (see section 3.2.5 below). Some animals experienced multiple GA procedures during this time – which may be multiple MRIs or both MRI and surgical interventions.

3.2.2 MRI

After exclusion criteria (see section 3.2.5 below), there were 35 MRI events from 14 different animals used for analysis. MRI procedures are often repeated, several weeks, months or years apart, as part of the standard neuroscience protocols. Therefore, some data collected for this research were obtained from a repeated MRI exposure in the same individual (Table 3.1). Data obtained may not be from an animal's first experience of MRI or GA, even if they only experienced one MRI within this dataset. Length of MRIs varied from 1 to 9 hours (mean and SD: 5 ± 1.9). The 14 macaques ranged in age from 4-8 years old (mean and SD: 5.7 ± 1.1) and weighed 7.9-14.9kgs (mean and SD: 10.6 ± 1.8).

Table 3.1: The number of macaques used in this study, split by the number of MRI events they experienced to contribute to the dataset. Overall, there were 35 MRIs from 14 different macaques after exclusion criteria were applied (see section 3.2.5).

Number of individuals	MRI datasets in this study
9	1
2	2
1	3
1	9
1	10

3.2.3 Headpost surgery

In total, 11 headpost surgeries (macaques = 11) were able to be included in the analysis following exclusion criteria (see section 3.2.5). Of these, 4 macaques were fitted with MRI compatible headposts and 7 with electrophysiology, titanium headposts (see section 2.4.7). Of the 11 macaques, 6 also experienced MRI events which feature within the MRI dataset. There is a minimum of 8 months between the MRI and headpost events. The duration of surgeries ranged from 4.5-7.5 hours (mean and SD: 5.7 ± 1.0). Macaques ranged from 4.8-8.6 years old (mean and SD: 6.0 ± 1.0) and weighed 8.4-13.8kgs (mean and SD: 10.9 ± 1.7).

3.2.4 Accelerometers and collars

Prior to both MRI and surgical interventions, macaques were fitted with a customised, soft neck collar containing a uniaxial, lightweight piezoelectric accelerometer (CamNtech Ltd Actiwatch®-mini). Accelerometers recorded activity levels (generic term denoting the amplitude of movement detected by the device) in either 10 or 30 second epochs (allowing continuous data collection for 15 or 45 days respectively). See section 2.1 for more details.

Collars were fitted around the neck of the animal whilst either under GA or under restraint, the primary purpose of which was for neuroscientific research. Collars were fitted by appropriately trained personnel approximately 2 weeks prior to an intervention and removed up to 2 weeks after. Collars stayed on continuously during this time period, with the exception of a refitting if required due to battery expiration. In this case, refitting was done in the same day.

Data were collected from repeat treatments in cases where individuals were exposed to more than one intervention during the study.

3.2.5 Data processing and analysis

Daily means of activity (between 07:00 and 19:00) were calculated from 10 or 30-second resolution data for each day of bio-logging and used for all subsequent analyses. Only those interventions for which data were recorded i) both before and after the event and/or ii) without data gaps following the intervention were used. When these criteria were not met, the whole event was removed from analysis. Following this exclusion criteria, 35 MRI events from 14 macaques and 11 surgery events from 11 macaques were used for analysis. Missing data could have resulted from delays with collar fittings, removal of collars prematurely or read errors in the data files resulting in corrupt actigraphy data. Baseline activity levels were estimated for all interventions by taking the mean and standard error of daytime activity over all days prior to events. Only events with 3 or more days of data prior to interventions were used to ensure incorporation of natural variability in activity (see 'Appendix 1' for details of randomization analyses that informed this method of selection).

All analyses were undertaken in R 3.0.2 (R Core Team, 2013). R functions *gam* and *gamm* from package *mgcv* were used for Generalised Additive Models (Wood, 2011). Package *stats* was used for functions *AIC* and *loess*, to generate Akaike Information Criterion and LOESS smoothers. (R Core Team, 2021). Generalised Linear Models were created using function *glmer* from the *lme4* package (Bates et al., 2015).

3.2.6 Assessing recovery

To determine the time taken to return to baseline activity ranges following interventions, the difference in daytime activity between baseline periods and mean daytime values on each day following an intervention was calculated. To investigate the time taken to return to baseline

activity ranges at the population level, a series of candidate threshold models were initially run on the full MRI dataset during methods development and compared using model comparison (Akaike Information Criterion, AIC – see 'Appendix 1' for details). In addition, Generalized Additive Mixed-Effects models (GAMMs) with a random intercept for 'individual' were fitted, with 'difference in activity from baseline' as a response and 'days since intervention' as a fixed effect. To account for potential biases that duration of anaesthesia may have on activity, data were constrained in these models to exclude very short and long GA events (<3 hours and >7.5 hours, n=7 sessions, leaving 28 MRI events from 12 macaques available for inclusion in this analysis). The mean recovery time of the population was estimated to be the time (in days) at which the 95% confidence band of the fitted values from the model intersected with an activity difference of zero. The same method was applied to activity data surrounding surgical interventions, although in this case Generalised Additive Models without a random effects structure were implemented, as there were no repeat measures across individuals.

To investigate the time taken to recover from each individual MRI and surgical event, LOESS smoothers were then fitted through data for each event (regardless of GA duration), and the time point (*t* in days since intervention) at which the LOESS curve entered the observed range of baseline activity values (defined by the standard error of activity during baseline periods) was used as an estimate of the recovery time (see Figure 7b for an example plot). Recovery time was assumed to be zero (i.e. no effect) in cases where activity differences never left the range defined by the standard errors (SE belt) of baseline values. In cases where activity differences never returned to the observed baseline range, the last day of data collection following the event was used as the recovery time.

To investigate the effect of various MRI factors on recovery time (time taken to return to baseline activity levels), Generalised Linear Models (GLMs) with a negative binominal error distribution and logit link function were used. Sample sizes were inadequate to run this same

analysis on surgical intervention data. Collinearity between candidate MRI variables (MRI duration, macaque age, macaque weight, number of previous MRI exposures, no of previous GA exposures, time since last MRI and time since last GA) was determined by calculating variance inflation factors (Zuur et al., 2009), and only non-collinear variables were used in models (terms for weight, number of previous GAs and time since last GA were dropped). Generalized Linear Mixed-effects Models (GLMMs) with a random intercept for individual, negative binomial error distribution and logit link function were initially run to account for dependence in observations from the same animal, however the level of between-group variability was zero and thus did not warrant inclusion of this term in the model. Model selection was performed using a forward stepwise selection method involving Akaike Information Criterion (AIC). The most parsimonious model was taken to be that which minimised the AIC, in cases when the difference in AIC between models was >2 (Burnham & Andersen, 2002). When this processes resulted in multiple models with an AIC of <2 (i.e. no evidence for a difference), a model averaging approach was taken and Akaike parameter weights calculated (Symonds & Moussalli, 2011).

A formal comparison of recovery time between MRIs and surgeries was undertaken using linear mixed-effects models with a random intercept for individual.

3.2.7 Modelling of cumulative effects

Datasets from 2 of 23 macaques were sufficient in size to allow a preliminary investigation of the potential cumulative effects on daytime activity of repeated exposure to MRI events. Generalized Additive Models (GAM) were applied individually to data from the two animals (number of MRIs, animal 1 = 9; animal 2 = 10). The difference in activity between baseline periods prior to MRIs and post-MRI periods of 6 days was fitted as a response, while 'time since first MRI' (Time) and 'duration of anaesthesia' (Dur) were fitted as fixed effects. A 6-day post period was adopted based on the obtained 'recovery time' estimate from population-level models. Prior to modelling, candidate fixed effects were tested for collinear following methods outlined above and

terms for ‘age’ and ‘number of previous MRI events’ were dropped. Model selection and validation processes were used as previously outlined (and following approaches in (Zuur et al., 2009)).

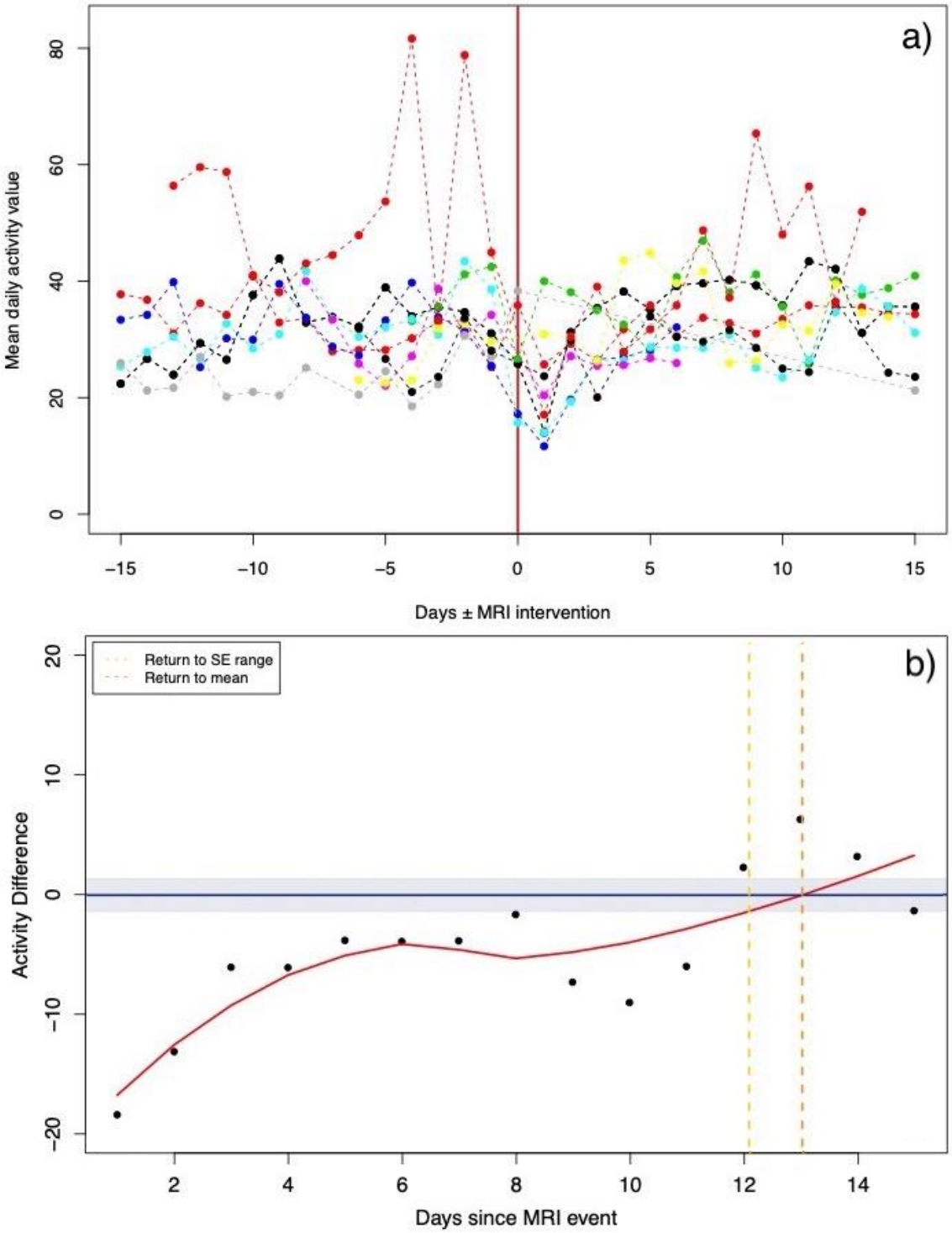

Figure 7: ***a)*** *Example traces of a group-housed male macaque before and after MRI interventions (vertical line = time of intervention). Each line represents a different MRI event for the same macaque. Graph* ***b)*** *shows the activity difference from baseline (absolute change) on each day post-MRI and is an example of the loess smoothing method used to estimate time taken to recover to baseline activity levels following an MRI (dotted yellow line = time at which the loess curve intersects with the baseline activity range defined by the SE in activity for pre-intervention data; dotted orange line = time at which loess curve intersects with an activity difference of zero).*

3.3 Results

Baseline activity values for each event varied both between and within individuals. For MRI procedures, baseline ranged from 10.73 to 52.46 (mean and SD: 29.5 ± 10.3) and for surgical procedures from 12.09 to 37.02 (mean and SD: 22.9 ± 7.9).

3.3.1 Effects of MRI interventions

Pre and post-event activity data for 35 MRIs across 14 animals were recorded during the study. For the GAMM, to avoid potential bias caused by the differing durations of anaesthesia, very short and very long GA events were excluded from this analysis (<3 hours and >7.5 hours, leaving 28 MRI events from 12 macaques available for inclusion in this analysis). There was a significant relationship between daytime activity and days since exposure to an MRI, although the deviance explained by the model was small (Table 3.2). On average, macaque activity returned within baseline ranges 5.7 days after an MRI intervention, following decreased levels of activity (Figure 8). Threshold models run during data exploration phases indicated negligable between-individual variance in the slope of the relationship between days since MRI intervention and activity (see 'Appendix 1' for results of piecewise regression models).

*Table 3.2: Coefficients from a GAMM with a random intercept for 'individual' run to model the relationship between days since MRI intervention and daytime activity in a captive population of group-housed Rhesus Macaques (n, individual = 12, intervention = 28 after exclusion of short (<3 hours) and long (>7.5 hours) GA events).*p<0.01, **p<0.001.*

Terms	**Estimate**	**Standard error**	***t*-value**	***n***
Intercept	1.779	0.617	2.882*	360
	edf	***F*-value**	**Adj. R^2**	
s(Days since MRI)	2.437	20.550**	0.119	
			Random intercept	**Residual**
		Variance	1.760	63.797

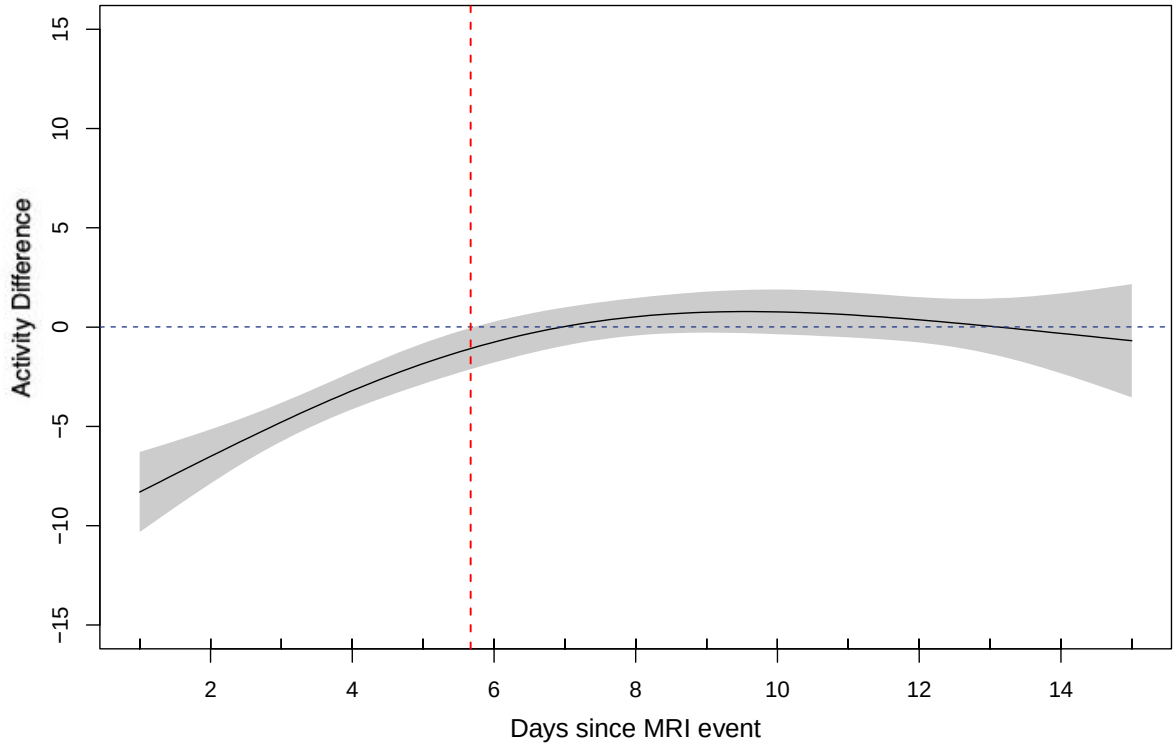

Figure 8: Fitted smoother and 95% confidence bands from a GAMM run to model the relationship between days since MRI intervention and difference in activity from baseline levels (absolute change) in a captive population of group-housed Rhesus' Macaques (n, individual = 12, intervention = 28 after exclusion of short (<3 hours) and long (>7.5 hours) GA events). The time taken to return to baseline activity levels (vertical dotted red line) at the population level was estimated as the point at which the 95% confidence band intersected with an activity difference of zero.

Time taken to recover to baseline levels ranged between 1 and 13 days for events when the animal moved out of the observed pre-intervention activity range following the MRI (mean and SD = 6 ± 3.9). Macaques did not leave their baseline activity range after 2 of 28 events, and in 5 cases had still not returned to pre-intervention levels by the time data collection ceased. Similarly, time taken to intersect with a difference in activity from baseline of zero ranged between 2 and 13 days, with 9 events during which macaques did not intersect with zero.

Despite some exogenous variability in daytime activity values, macaques exhibited an decrease in activity following most MRI interventions (92%), with only two clear cases of an increase in activty.

The best models for the effect of MRI variables on recovery time included only duration of anaesthesia and number of previous MRI events as fixed effects, and indicated that moderate duration interventions (>3 - ≤6 hrs) had a greater effect on recovery time than short duration interventions (≤3 hrs; Figure 9 and Table 3.3). Netherlessless, the most parsimonious model only explained 9.6% of deviance in the data and was not significantly different from the null model (LRT, R^2= 4.055, *P* = 0.132; ΔAIC of null model compared to final model = +0.06). 'Animal age' and 'time since previous MRI intervention' did not improve model fit using an ΔAIC criterion of >2 (Table 3.3).

*Table 3.3: Results of model selection for GLMs with a negative binomial error distribution and logit link function, run to test the relationship between various MRI-related factors (duration of MRI anaesthesia = Dur; age of individual = Age, no. of previous MRI interventions = nPrev; time since last MRI intervention = tPrev) and recovery time (time taken to return to baseline activity levels; TRec) in macaques (n, individuals = 14, MRI interventions = 35). Candidate models with measures of Akaike Information Criteria (AIC), delta AIC components (ΔAIC) and Akaike weights (AIC wt) are shown. The model with the lowest AIC taken at each stage of forwards stepwise model selection is shown, as are coefficients from the model parsimonious model. The three best models are presented in bold. **p<0.05*

Model	**AIC**	**ΔAIC**	**AIC wt**
TRec ~ 1	**203.4**	**0.055**	**0.309**
TRec ~ Dur	**203.3**	**0.000**	**0.318**
TRec ~ Dur + nPrev	**203.6**	**0.332**	**0.269**
TRec ~ Dur + nPrev + tPrev	205.6	2.317	0.100
TRec ~ Dur + nPrev + tPrev + Age	212.5	9.224	0.003
TRec ~ Dur coefficients	**Estimate**	**Standard error**	***z*-value**
Intercept	1.011	0.426	2.372**
Duration (>3 - ≤6hrs)	0.934	0.452	2.069**
Duration (≥6hrs)	0.738	0.499	1.478

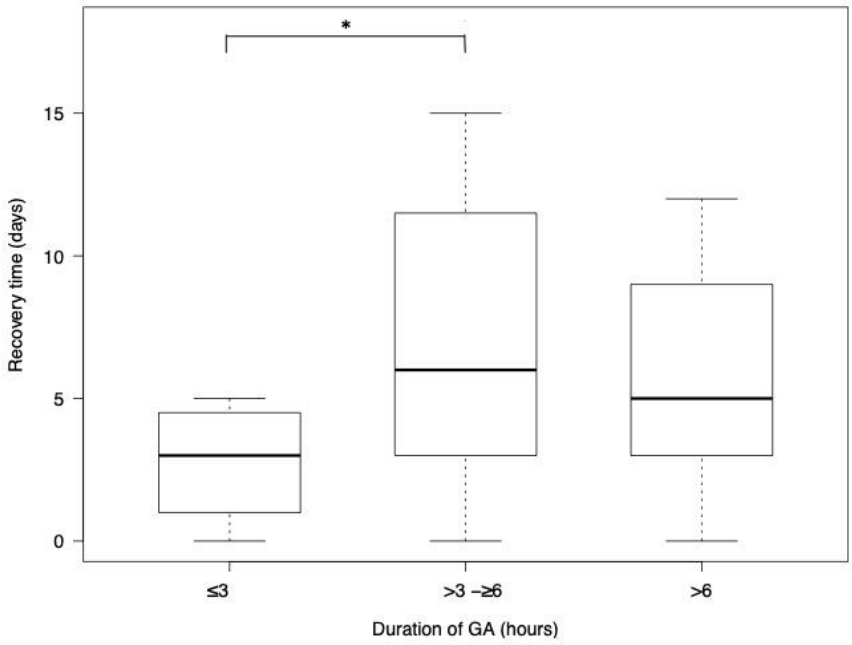

*Figure 9: Boxplot for the effect of duration of MRI anaesthesia (hours) on recovery time (time taken to return to baseline activity levels) in macaques (*n*, individuals = 14, MRI events = 35 (≤3 = 4, >3-≥6 = 23, <6 = 8)). Asterisks indicate significant differences between duration factor levels as indicated by a negative binomial Generalized Linear Model (where* P *< 0.01; see Table 3.3 for parameter estimates).*

3.3.2 Effects of surgical interventions

Activity data for 11 surgical interventions, from 11 animals, were used to test the influence of this form of treatment on activity. As with MRI events, a significant relationship between daytime activity and days since exposure to surgery was found (Table 3.4), with macaque activity intersecting with baseline ranges at an average of 2.6 days following surgical intervention (Figure 10). Again, this model only explained a small amount of deviance in the data (14.5%).

*Table 3.4: Coefficients from a GAM run to model the relationship between days since surgical intervention and daytime activity in macaques (n, individual = 11, intervention = 11). *p<0.001.*

Terms	**Estimate**	**Standard error**	***t*-value**	***n***
Intercept	-2.579	0.810	-3.186*	171
	edf	***F*-value**	**Adj. R^2**	**Deviance**
s(Days since surgery)	3	9.435*	0.130	14.5%

At the individual level, time taken to return to baseline levels ranged between 1 and 9 days following surgical intervention (mean and SD = 4 ± 2.8). All macaques left their baseline activity range after surgery. However, in 3 cases they had not returned to pre-intervention levels by the time data collection ceased. Despite some variability in the data, most macaques (91%) showed decreased activity in the first few days following surgery, and in 7 cases then engaged in increased levels of activty that extended beyond baseline levels and remained elevated for a number of days (Figure 10). This activity increase contrasted with the pattern seen in animals after most MRI events, which remained close to baseline levels after first returning within their pre-intervention range .

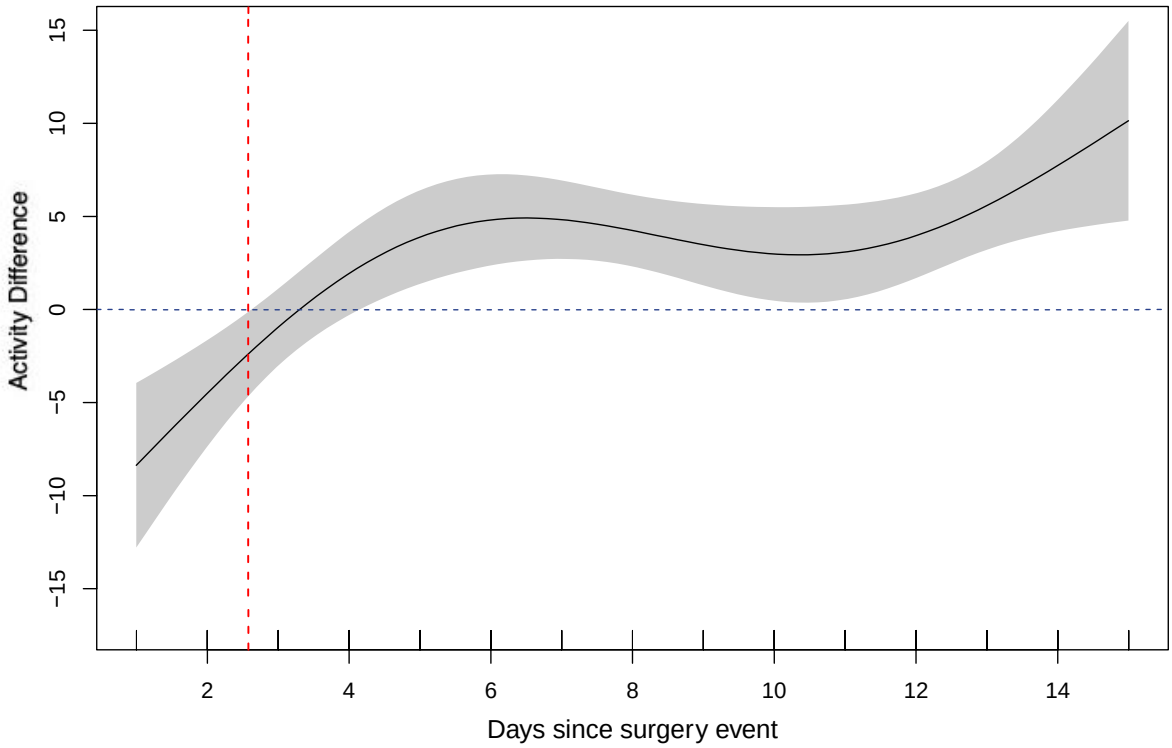

Figure 10: Fitted smoother and 95% confidence bands from a GAM run to model the relationship between days since surgical intervention and difference in activity from baseline levels (absolute change) in macaques (n, individual = 11, interventions = 11). The time taken to return to baseline activity levels (vertical dotted red line) at the population level was estimated as the point at which the 95% confidence band intersected with an activity difference of zero.

When directly comparing 'recovery time' between MRI and surgical interventions (Figure 11, for events with GA durations constrained to a range of ≥3 and ≤7.5 hours), no significant difference was found between the two treatments (LRT, R^2= 0.052, *P* = 0.820; ΔAIC of null model compared to model with treatment as a fixed effect= +0.97). Recovery time reflected here includes only time to return to baseline activity and does not account for the subsequent increase in activity seen post-surgery.

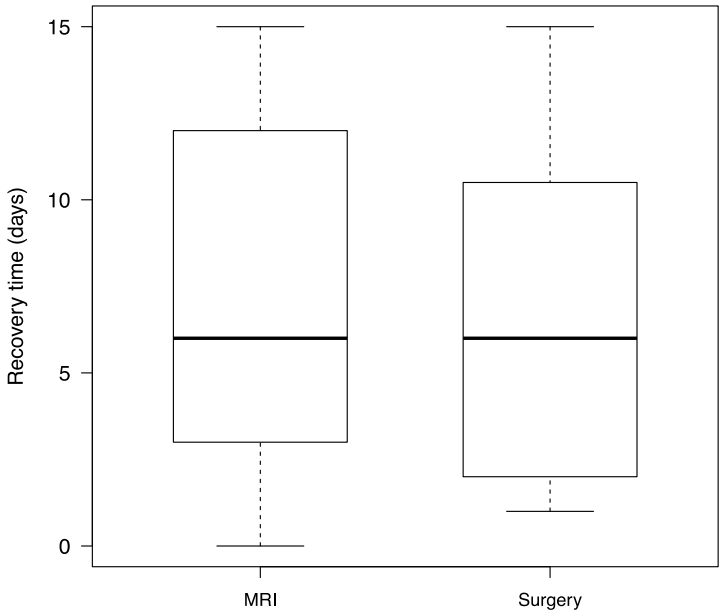

*Figure 11: Boxplot for the effect of treatment (MRI or surgical intervention) on recovery time (time taken to return to baseline activity levels) in macaques (*n*, individuals = 18, MRIs = 28, surgeries = 11 when constrained to events with GA duration of ≥3 and ≤7.5 hours).*

3.3.3 Cumulative effects

GAM was used to model the impact of multiple MRI exposures on differences in activity levels following an MRI event. Two macaques experienced several MRIs under GA over 3 years (animal 1 = 9; animal 2 = 10). The best models for both animals were those that included both time since first MRI (used to represent the accumulation of MRI events) and duration of GA as fixed effects (Table 3.5; Figure 12). Nevertheless, for animal 1, there was no evidence to suggest that the model including these fixed effects had more predictive power than the null model that

only contained the intercept. In contrast, for animal 2, both fixed effects contributed significantly to the deviance in the data (Table 3.5; Figure 12).

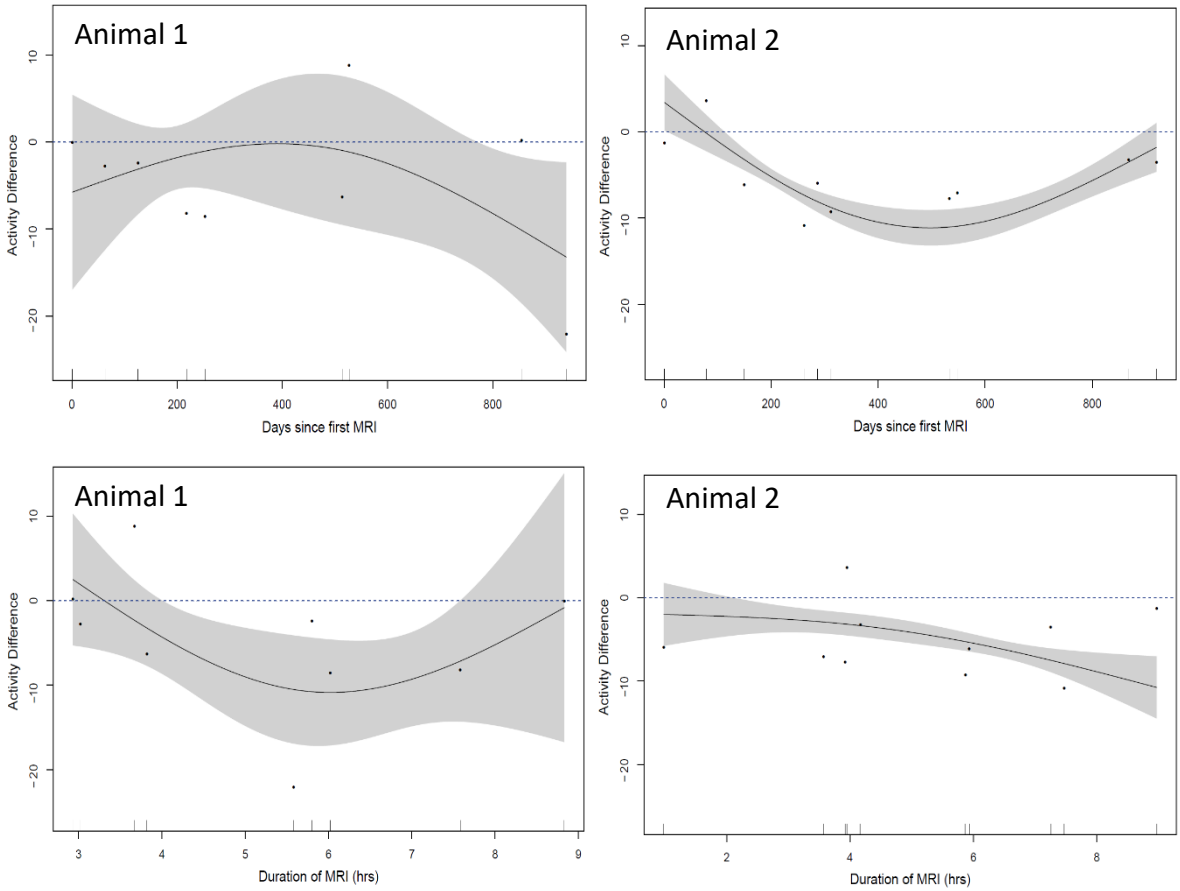

Figure 12: Fitted smoothers and 95% confidence bands from individual-level GAMs Models run to test the cumulative influence of MRI exposure on daytime activity patterns in two macaques (animal 1: n = 9; animal 2: n = 10). Activity difference = difference in activity between baseline periods (absolute change) prior to MRIs and post-MRI periods of 6 days.

*Table 3.5: Results of model selection for GAMs run to test the cumulative influence of MRI exposure on daytime activity patterns in macaques. Models were applied individually to data from 2 animals that had been exposed to multiple MRIs (animal 1: n = 9; animal 2: n = 10). The difference in activity between baseline periods (absolute change) prior to MRIs and post-MRI periods of 6 days (ActDiff) was fitted as a response, while time since first MRI (Time) and duration of anaesthesia (Dur) were fitted as fixed effects. Candidate models with measures of Akaike Information Criteria (AIC), delta AIC components (ΔAIC) and Akaike weights (AIC wt) are shown. The model with the lowest AIC taken at each stage of forwards stepwise model selection is shown. For each animal, parameter estimates are shown from the best model. *p<0.05, **p<0.01, ***p<0.001.*

	Animal 1			**Animal 2**		
Model	**AIC**	**ΔAIC**	**AIC wt**	**AIC**	**ΔAIC**	**AIC wt**
ActDiff ~ 1	66.897	0.694	0.29	60.124	12.597	0.002
ActDiff ~ Time	67.872	1.669	0.18	61.904	14.376	0.001
ActDiff ~ s(Time)	68.695	2.493	0.12	55.371	7.843	0.019
ActDiff ~ s(Time) + s(Dur)	66.203	0.000	**0.41**	47.528	0.000	**0.978**
Coefficients	**Estimate**	**SE**	***t*-value**	**Estimate**	**SE**	***t*-value**
Intercept	4.578	2.458	1.863	5.128	0.639	8.020***
	***F*-value**	**Adj. R^2**	**Deviance**	***F*-value**	**Adj. R^2**	**Deviance**
s(Time,2)	1.423	0.239	62%	17.009**	0.771	87%
s(Dur,2)	2.115			5.671*		

3.4 Discussion

This study investigated the immediate and cumulative effects of GA on the activity levels of captive-housed macaques. Accelerometers provided a quantitative and minimally invasive approach to explore changes in the physical activity of freely moving animals. A decline in activity is observed immediately following both MRI and surgical procedures. There was some evidence to suggest that the duration and number of previous exposures to GA influences the time taken to recover pre-GA activity levels. However, there appears to be a lot of variation between individual macaques.

3.4.1 Implementing novel, non-invasive methods

The data collection protocols established and implemented as part of this project provided a simple and effective technique for assessing the physical activity of laboratory-housed rhesus macaques. Using customised, soft, dog collars to house accelerometers appeared to cause minimal initial discomfort to the animals, following several hours of habituation. Based on behavioural observations, wearing the collar did not appear to influence either animal welfare or impact the primary research objective. This method allowed for individuals to continuously wear the collar throughout the data collection period, with the exception of a short removal period to change or reset the actiwatch, when storage capacity was reached. Therefore, following initial collar fitting, there was no further disruption to the animals during the data collection period and so the methods themselves are unlikely to influence activity data.

Additional benefits of implementing this actigraphy method were that the collars were easy to fit and accelerometers simple to set-up. As the collars were fitted under restraint, it was also possible to train some neuroscience researchers to fit the collars safely, allowing fittings to work around their experimental schedules without interference and minimising the time and resources

required to obtain the data. Using accelerometers allowed for continuous, quantitative measures of physical activity throughout the day and night.

Obtaining actigraphy data prior to the GA procedures provided baseline activity for each individual event. This is important as there was variation between baseline periods of activity both within and between individuals. These variations in activity levels could be caused by multiple factors, such as the age of the animal and their social housing situation. Such factors can change overtime, so when taking multiple measures from the same animal, as during the MRI study, it was important to re-establish the baseline. Furthermore, if collars were fitted after the GA procedure, then the activity levels observed following recovery could partially be influenced by re-habituation to the collar.

3.4.2 Administration of GA: MRI vs surgery

The disruption that GA inflicts on the bodies physiological functions suggests that its administration is a stressor. It is widely acknowledged that there is a link between stress and physical activity, with most evidence suggesting that stress impedes physical activity (Stults-Kolehmainen & Sinha, 2014). Therefore, a decline in activity levels is to be expected following a stressful event. This is observed in the macaques following both GA procedures, where activity levels decline from baseline, with an average recovery time of 5.7 days or 2.6 days for MRI and surgical procedures, respectively. As it can take some time for anaesthetic drugs to clear the system, some decline in function, including a reduction in activity levels, is anticipated in the first few hours immediately following a GA. However, the impact on some physiological parameters may suffer longer-term disruption. Following human surgical procedures, there is evidence of patients taking several days or weeks to return to pre-surgical physical activity levels, dependent on factors such as the type of surgery and demographic of the patient (Bisgaard et al., 2002; Ratcliffe et al., 2021; van der Meij et al., 2017; Wasowicz-Kemps et al., 2009). In rats, there is

evidence of a decline in locomotor activity for several days following a minor surgical procedure, involving a 30-minute GA (Hendrickx et al., 1984).

The recovery period following a surgical procedure does often require individuals to be less mobile, to reduce post-surgical pain and allow wound healing. Therefore, from the human literature, it is difficult to determine whether these longer recovery periods are attributed to the GA, the tissue trauma or both. In humans, exposure to GA is almost always done as part of a surgical procedure, so potential effects of surgery or GA cannot be determined. However, these studies have also shown variation in activity recovery depending on the severity of the surgical intervention. For example, in a comparative study in patients undertaking either 'minor' or 'intermediate' surgeries, 'minor' surgeries showed a faster recovery rate, with pre-op activity levels regained after 3 weeks, compared to more than 5 weeks for 'intermediate' surgeries (van der Meij et al., 2017). Furthermore, activity levels after laparoscopic surgeries in humans showed just 2-3 days was required to return to baseline (Bisgaard et al., 2002). A comparative study of gastrectomy methods showed 2-3 days was required to achieve pre-intervention activity levels following laparoscopic surgeries but a longer period of more than 7 days was required following open surgery (Fujita et al., 2003). In dogs, significant activity change was observed following open but not laparoscopic ovariectomies (Culp et al., 2009). This implies that the severity of the procedure conducted has an impact on the change to activity levels and the time needed for recovery, with a surgery with higher severity requiring a longer recovery period. Interestingly, in terms of time to return to baseline activity levels, we have not seen this pattern in the macaques, with the GA involving no tissue trauma (MRI) taking a longer 5.7 days to see a return to baseline activity levels, compared to 2.6 days for headpost surgery. Furthermore, direct comparison of the time to recovery (days) between MRI and surgical intervention (Figure 11), revealed there was no difference between the two procedures. It could also be the case that a return to baseline activity levels does not indicate the end of recovery, especially if the animal does not then maintain baseline activity, as seen in the post-surgical dataset of this study.

The apparent speedier return to baseline activity of macaques following surgical intervention, compared to MRI, could be attributed to numerous factors. Firstly, the best fitting models in this research only account for a small amount of the deviance in the data, 9.6% and 14.5% for MRI and surgical GAs respectively. Therefore, a lot of what influences an animal's time to return to baseline activity levels is still unexplained. Variation between individuals is large, which suggests that the influence of GA is not a standardised response but that each individual animal experiences recovery differently. Individualised response to stressors is common when studying welfare parameters, as individuals commonly adopt different coping mechanisms or are affected in a different way (Broom & Johnson, 1993). For example, startle responses to the same stimuli within a species can vary from freezing behaviour to attempts to flee. Individualised responses also account for why we might see a difference in recovery between the procedures when treating the data at the population level, as with the GAM's (Figure 8 and Figure 10), but not when we do a direct comparison of each individual's recovery time using GLM (Figure 11). This emphasis on individualised responses also suggests that the results from the MRI and surgical datasets are not completely comparable, as the data from each comes from a different set of animals, with seven individuals appearing in both datasets.

Another possible explanation for speedier recovery might be that experiencing multiple MRI events is influencing the recovery. Unlike the headpost surgery, MRI procedures often occur multiple times throughout a macaque's neuroscientific career, and this was reflected in some of the data we were able to obtain. The repetition of a procedure in an individual offers opportunity for that individual to learn or become habituated to the GA protocol. A study in horses showed that multiple exposure to GA for MRI resulted in an increased quality of recovery with experience, with implication that they achieve this by lengthening their recovery duration (Platt et al., 2018). However, this study only reflected short term recovery, over several hours, rather than a difference of several days. It does seem valid that extending your recovery period would offer a better-quality recovery overall. Following human surgeries, return to physical activity levels often

takes several weeks, which could be attributed to advise from physicians for patients to 'take it easy' following surgery to encourage healing.

Following the return to baseline activity levels, animals who experienced an MRI event then maintained activity levels and surgical subjects increased physical activity (Figure 10). One suggestion for this increase might be an increase of restless or agitated behaviours, such as pacing. Increased restlessness has been acknowledged as an indicator of increased pain, for example when castrating lambs without local anaesthetic (Molony et al., 1993). Restlessness could indicate ongoing pain or stress caused by the surgical procedure or could be a consequence of the change to the animal's daily routine. Often, to allow recovery and wound healing, post-surgical macaques have a break from their daily testing routines for several days or weeks, meaning they are not removed from the home cage to participate in cognitive tasks as they would have been prior to the surgery. This is not usually the case following an MRI, where animals usually resume testing within 1-2 days. This may cause an increase in daily activity levels, as the testing routines require restraint, therefore ensuring the immobility of the animal for a short period of the day. Alternatively, increased activity could reflect the animals interfering with the accessible sutures and wound following surgery. Animals are prone to picking at wound edges and sutures; this is observed in macaques following headpost surgeries (Perry et al., 2021). Due to the placement of the collar around the neck, frequent movement of the head to facilitate wound interference or scratching around the area could be easily detected by the accelerometers and thus drive overall activity level up. To establish the source of increased activity, a future study could explore activity variance and whether the overall daily increase is due to bouts of high intensity activity, suggesting restlessness, or frequent low-level movements, suggesting regular wound picking. As activity levels remained elevated throughout the 15-day monitoring period, it would also be beneficial to monitor for longer periods or measure activity of individuals again several weeks or months following the surgery, to establish how prolonged this change was.

3.4.3 Length of and previous exposure to GA

Duration of the anaesthetic and the number of previous MRI events an individual macaque experiences may have some influence on the time required to return to baseline activity following a GA for MRI. The inclusion of GA duration and previous MRI number as fixed effects when modelling the time taken for recovery did not significantly improve the fit, compared to the null model, but did show very similar model outputs (see Table 3.3). If these factors do effect recovery time, the effect is likely to be small. Due to the influence of individual macaque variation on recovery time, it is also likely that the best fitting model, containing one, none or all of these effects, might vary depending on the individual. Adding animal age and time since previous MRI did not improve the null model, suggesting they influence no effect on recovery time. However, the age range was narrow (4-8 years) and the time periods between MRI events were often several months, with the minimum being 13 days. Therefore, the effect may be undetectable within the ranges of this study.

A direct comparison of MRI durations revealed a significantly longer recovery period following an MRI of moderate length (>3-≤6hrs) compared to one of short duration (≤3 hours). Longer exposure to anaesthetic agents results in longer disruption to the many physiological processes, increasing the risk of complications and the potential time needed for recovery. Duration of GA is considered a risk factor for poor quality or prolonged recovery in humans and other animals, such as horses (Loomes & Louro, 2021; Sinclair & Faleiro, 2006). However, studies reporting this effect are referring to a prolonged emergence times, i.e. time to recover from GA in minutes or hours. No evidence within the literature could be found on GA duration and longer term aspects of recovery, such as returning to previous activity levels. Interestingly, this study found no difference in recovery time following MRIs of a long duration (>6 hours). Due to the opportunistic data collection method used in this study, there was no control over the duration of the MRIs from which data were collected. Therefore, the comparison of durations had an

unbalanced design, with many more MRIs falling to the 'moderate length' category (n=23), compared to the short (n=4) and long (n=8) durations. However, although not statistically significant, the models from this study do highlight that there is a possible impact of GA duration on recovery time. Further investigation to confirm this finding could lead to potential protocol refinements to improve animal welfare, such as conducting multiple, shorter GA events in place of one of longer duration.

The MRI data for this study contained frequent, repeated exposure to MRI from two animals over a 2.5-year period. The dataset from these two animals was used to further investigate the effect of anaesthesia duration and the time since their first MRI event on each animal's activity difference from baseline, 6 days post-GA. In one of the two macaques (Animal 2), both of these factors had a significant impact on the return to baseline activity. As these two animals shared a similar experience, this further highlights the individual variation in response to GA. In the case of animal 2, a longer anaesthetic seems to prolong the recovery time of the macaque, which aligns with what was expected and observed previously. However, the number of previous MRIs (represented as time since first MRI) attributed a greater influence over activity levels. Animal 2 showed a greater decline in activity levels from baseline with increasing number of MRI events initially, followed by a lessening difference after 6-8 MRIs. It is worth noting that this lessening difference appears to be driven by the final two MRI events. The baseline for these MRI events is much lower than all other MRI events from this animal, with the baseline of the first 8 MRIs ranging from 35 to 43, compared to 13 and 11 for MRI 9 and 10, respectively. Difference in activity for each MRI event was calculated using a new baseline value for the time period immediately preceding the event. Therefore, it is likely that difference in activity value is much lower for these events because baseline was much lower to begin with. Interestingly, prior to the final 2 events, this animal experienced a change to social grouping, going from being housed with 4 other macaques to being singly housed following social conflicts. This could be a potential driver for the decline in activity levels. Behavioural differences are observed between singly and pair or

group-housed macaques and behavioural disruption is seen during periods of social transition (Hannibal et al., 2017).

In a review to assess the cumulative and lifetime experience of NHPs used in neuroscience research, the animal procedures committee outlined several scenarios to describe how repeated exposure to neuroscience procedures could influence animal welfare (Pickard, 2013). In the case of animal 1, the data reflected the proposed scenario where full recovery occurs between MRI exposures and there is no additive effect on the animal's activity recovery. However, this scenario does not apply to the data observed from animal 2, where there does seem to be cumulative effect. The alternate scenarios proposed in the review would better fit this pattern. These scenarios describe either partial or no recovery of the welfare compromise inflicted by the procedure, allowing subsequent procedures to have an additive effect on the welfare state. This would cause welfare to be further compromised and, in this case, leads to a change in the recovery of activity levels post-GA. This change in recovered activity could also be the result of habituation to the GA procedure, as mentioned previously.

There are additional potential factors that could influence the recovery of the animal which we were unable to analyse during this study. For example, whether the individual is socially or singly housed or their dominance rank. Although these data are available for the animals and procedures used in this study, the analysis is not viable due to unbalanced sample sizes for comparison. This is due to the nature of this project, where data were collected opportunistically from any available animal and parameters could not be altered, to avoid further welfare impact or influence on the neuroscientific studies.

3.4.4 Conclusions

Actigraphy is a reliable tool for assessing the physical activity levels of NHPs. The findings from this study show a decline in macaque physical activity, lasting several days, following GA procedures. This may imply a decline in welfare state and highlight the need for potential

refinements which may minimise this impact or prevent further welfare decline. Furthermore, a surgical procedure is also accompanied by an increase in activity levels following the initial decline, suggesting behavioural change. GA duration and number of previous GA events may influence the time taken to recover from a GA. However, individual variation affects a large influence on activity levels.

This study has identified an objective, quantifiable measure of welfare that can be applied alongside existing scientific protocols to gain an understanding of NHP welfare. It is non-invasive, simple and effective at detecting behavioural change. Monitoring these changes offers valuable insight into the welfare impact of potentially stressful experiences. Additionally, this method is suitable for longitudinal welfare assessment, presenting opportunities to gain an understanding of cumulative experience.

Chapter 4 - Automated behavioural assessment using accelerometers

4.1 Introduction

Animal behaviour is a key component of animal welfare assessment. It is a readily accessible and commonly used welfare indicator, providing an indicator of overall health and reflecting the expression of an animal's own decisions (Dawkins, 2003). Behavioural-based measures have also been used to assess psychological health. Preference testing and monitoring stereotypies are just a few of the approaches available for gaining insight into the affective state of animals (Dawkins, 2003). Furthermore, natural living and, concurrently, the presence of natural behaviours is thought to play an essential role in the welfare of captive animals (Fraser, 2008). Therefore, obtaining behavioural data is essential for an overall picture of animal welfare.

Monitoring the behaviour of animal's used in research is a key component to their ongoing welfare assessment. For example, identifying natural behaviours speaks to the suitability of husbandry conditions and behavioural indicators of stress or suffering can help to highlight intervention points (Hubrecht, 2014). A report from a joint working group (BVAAWF/FRAME/RSPCA/UFAW) on refinements for NHPs lists behavioural assessment as an essential component for assessing and facilitating primate welfare (Jennings et al., 2009).

The traditional method used for the assessment of animal behaviour is via direct observations. Qualitative measures of behaviour, such as narrative accounts, provide extensive detail and can be very insightful for the understanding of an animal's welfare (Fraser, 2009). A quantitative approach can be generalised to a larger population. Therefore, it usually provides the more repeatable, reliable evidence required to draw conclusions and implement changes to improve animal welfare. Behavioural observations are often quantified for this reason, for example by measuring the frequency or duration of a particular behaviour (Martin & Bateson,

2007). Whilst direct observations are invaluable, this method is not without its flaws. Observing animal behaviour is often time-consuming and may be limited by several factors such as the visibility of the animal, observer bias, and disturbance caused by observer presence (Brown et al., 2013).

To obtain a quantifiable measure of animal behaviour, techniques for monitoring behaviour are moving beyond direct observations and becoming more automated. Devices that can be temporarily attached to animals, such as accelerometers, can provide continuous and objective monitoring of individuals. Accelerometers have been used in many animal studies, providing an objective and reliable measure of movement through automated, continuous sampling. They have been used to compare locomotor activities between domestic species (Piccione, 2010), as a means to understand stress in shelter-housed dogs (Jones et al., 2014) and to document inactivity in response to pain in hens with keel-bone fractures (Casey-Trott & Widowski, 2018). A previous chapter in this thesis, 'Chapter 3 -The impact of general anaesthesia on non-human primates,' used accelerometers to monitor change in activity levels as a means to understand the welfare impact of anaesthetics.

Identifying patterns in actigraphy data (data from accelerometers) offers additional value for welfare research, as it can provide predictions of a range of animal behaviours. Actigraphy has been able to identify different types of locomotion, such as walking, trotting, and running, in cats (Watanabe et al., 2005) and cattle (de Passille et al., 2010). The ability to distinguish feeding, drinking, and resting behaviours from actigraphy data have been extensively reported in various farm animal species (Chapa et al., 2020; Watanabe et al., 2008).

Accelerometers have been used in several non-human primate (NHP) species to assess simple activity levels (marmosets (Mann et al., 2005), rhesus macaques (Sullivan et al., 2006) and vervet monkeys (McFarland et al., 2013)). Further behavioural distinctions have been made between locomotor behaviour in lemurs (Sellers & Crompton, 2004; Wunderlich et al., 2014) and

between whole body movements compared to arm or neck movements in rhesus macaques (Papailiou et al., 2008). Due to their complex behavioural repertoires, further classification of primate behaviours can be challenging. However, some studies have recorded success in identifying smaller, more subtle movements from accelerometer outputs which can be attributed to grooming or foraging behaviours (in chacma baboons (Fehlmann et al., 2017) and Japanese macaques (Sha et al., 2017)).

Behavioural indices of good welfare could be considered as a combination of the presence of normal behaviours and the absence of abnormal behaviours (Mellor et al., 2009). What constitutes normal behaviour in captive species is not always easy to define, particularly for those housed in laboratories. However, well-being is often associated with the ability to engage in species-typical behaviours (Novak & Suomi, 1989). In macaques, this may include social interactions, foraging, and grooming (NC3Rs, 2015). Abnormal behaviours are usually easier to identify and one such behaviour often discussed in the context of animal welfare is the presence of stereotypies. Stereotypies are repetitive, apparently functionless behaviours (Mason & Latham, 2004). Although not inherently an indicator of poor welfare, stereotypies are usually associated with a sub-optimal environment and may arise from central nervous system dysfunction, frustration, or repeated attempts to cope (Gottlieb, Capitanio, et al., 2013). Furthermore, stereotypies are the most commonly observed abnormal behaviour in rhesus macaques and pacing the most frequently seen amongst them (Lutz et al., 2003). Being able to identify both normal and abnormal NHP behaviours will allow for a more balanced view of overall animal welfare.

In this study, I am going to discuss the development of a 'rule-based model' based on behavioural observations and actigraphy data collected simultaneously from macaques used in neuroscience research. This model will make behavioural predictions from actigraphy data based on a set of rules which were created by identifying patterns in the actigraphy data that

corresponded to observed behaviours. The model will assign every activity epoch a behavioural category and thus predict the amount of engagement in species-typical behaviours. The different behaviours and the accuracy of predictions will be discussed. Furthermore, the intra- and inter-rater variability of behavioural observations was checked to identify potential observation bias. Case studies are presented to indicate how the model could be used in practice to assess NHP behaviour and welfare.

4.2 Methods

Actigraphy data and behaviour observations from time-matched video footage were compared and explored to define patterns that could identify specific macaque behaviours. During a pilot study, 5 species-typical behavioural categories were established as being detectable in actigraphy data (Table 4.1). This study was conducted by a DPhil student (Natasha Gillies) undertaking a 3-month rotation within the lab in 2017 and involved a collaboration with Professor Maarten De Vos from the Department of Engineering Science, University of Oxford. This then led to the creation of a rule-based model, which developed a set of rules to assign actigraphy epochs to 1 of the 5 behavioural categories.

Table 4.1: Ethogram of 5 species-typical behavioural categories.

Behaviour	**Code**	**Definition**
Inactive	IA	No movement but can be alert and vigilant
Active Rest	AR	Stationary with low-level movement of limbs. Includes self-directed behaviours, manipulation of objects, foraging, and grooming.
General Locomotion	GL	Low energy locomotion e.g. walking and climbing
Rapid Movement	R	Fast-paced locomotion including jumping, running, bouncing, & head shakes.
Pacing	P	Distinct, unchanging pattern of repetitive locomotion.

4.2.1 Accelerometers and collars

Rhesus macaques were fitted with soft, dog collars containing uniaxial, piezoelectric accelerometers (CamNtech Ltd Actiwatch® mini). Further description of collars, accelerometers and fitting protocols can be found in section 2.1. Accelerometers recorded activity levels at 10 second epochs. Information of housing and husbandry of the animals can be found in section 2.4.1.

4.2.2 Behavioural observations and cameras

Video footage was obtained from 12 adult macaques (5-10 years old, 7.4-14.5kgs weight, 1 female, 11 male) wearing accelerometers. The majority of subjects were pair-housed, with the exception of 3 which were group-housed (in the same social group) and 1 singly-housed. During video monitoring, all macaques continued to engage in neuroscience experiments, undertaking daily testing involving restraint and temporary (~1 hour) removal from the home cage. Up to 5 Sentient Pro 720p HDA security cameras were used in an attempt to capture the entire home cage of each animal. A digital video recorder (DVR) was used for video storage and observation.

To ensure the scored behaviours time-matched the actiwatch data exactly, the time on the DVR and the time on the laptop used to set up the actiwatch were matched (~1 sec) before every recording. Additionally, a 'shake test' was used to calibrate the datasets. This involved the actiwatches being vigorously shaken for 10 and 5 seconds at specific time points. This was done prior to collar fitting and following collar removal, providing a reference within the actiwatch data to time-match the DVR.

Video analysis was performed by researchers (RH, KC, NG) with experience of behavioural observations and an understanding of the typical behaviours of rhesus macaques. A single behaviour was assigned to every 10 second time window, to correspond with the 10 second length of an actiwatch epoch. Behavioural observations aligned with definitions from the

behavioural ethogram (Table 4.1). In instances where multiple behaviours were observed within a 10 second epoch, scorers recorded the primary behaviour. In most instances, this was the behaviour in which the majority of that epoch was spent. In epochs where 'rapid movement' was observed, this was recorded as the primary behaviour. This is because the nature of these types of behaviour are brief (often only a few seconds) but uncommon, so it would be of particular interest to animal welfare to ensure they were not missed.

Some video footage was analysed multiple times: by different observers or the same observer (or both), to obtain a record of intra- and inter-rater variability. In instances when one observer scored the same video multiple times, a time window of at least 2 weeks between observations was adhered to.

4.2.3 Creating a rule-based model

Behavioural observations from the video footage were used for model creation and further refinement to improve accuracy. A portion of observations were left out of model development for the purpose of validating the accuracy of the model.

The model was created by exploring corresponding actigraphy data and behavioural observations to identify ranges and patterns in the activity levels of each behaviour. Model building began by plotting the actigraphy ranges for each behaviour and simply assigning behaviours based on these raw amplitudes. Following this initial step, patterns were identified in the activity data, and these were used to create the rules which would predict the primary behaviour represented in each 10 second epoch.

A set of 6 model rules were created to assign each 10 second epoch of actigraphy data to 1 of the 5 behavioural categories. Activity level ranges significantly overlapped between behaviours, so behavioural classifications were initially grouped. The first model rule assigned behaviours to 1 out of 7 behavioural categories (see Appendix 2). For example, an activity value of

0 would be assigned to a category that represents either an IA or AR behaviour at this stage. The subsequent rules work to separate out these grouped categories from 7 down to 5 categories (see Table 4.1), based on factors including the length, mean, or standard deviation of behavioural bouts and differences between preceding epochs. The full set of rules can be found in 'Appendix 2.' During development, the rules were adjusted and modified several times, using new behavioural observations to check accuracy and provide further information. The final set of rules achieved the greatest accuracy of behavioural predictions. For further information on model development and the final rules, see 'Appendix 2'.

With the exception of the pilot study work, I was solely responsible for the model building, rule refinement, and testing of the model. I conducted 37.5 hours of behavioural observations that contributed to the development of this model (some of this is multiple observations of the same footage to test intra-rater variability).

Due to the nature of the 'piggybacking' approach to this research, actigraphy data were collected around many events, whenever opportunity arose, to assess the likely impact on animal welfare. For example, collars would be fitted to capture experimental procedures, such as a general anaesthetic, and husbandry related events, such as a relocation to a new room. Furthermore, accelerometers could also be deployed if requested by individuals working with the animals (researchers and animal care staff), as a means to assess a particular question. Such datasets can be run through the rule-based model and behavioural predictions observed and compared in relation to the specified event. Some of examples of this are presented in the form of case studies to illustrate the usefulness of the model in animal welfare monitoring and assessment.

Model development, testing, and figures were created in R (version 4.1.0) with RStudio (R Core Team, 2021). Plots were created using *ggplot2* (Wickham, 2016). Confusion matrix was generated using function *confusionmatrix* from the package *caret* (Kuhn, 2022).

4.3 Results

4.3.1 Rule-based model development

4.3.1.1 Behavioural observations

In total, 34.75 hours of video footage from 12 different macaques was scored for animal behaviours. Breakdown of how these hours were split for model development can be found in Table 4.2. Different footage from the same macaques was sometimes used in different areas of development, so total number of macaques reflects the number of unique individuals used.

Table 4.2: Breakdown of the hours of video footage observed to define behavioural categories, build the model, refine the rules and validate the final model. Macaques were all male, with the exception of one female used during model building.

	Defining behaviours & Model building	**Rule refinement**	**Validation**	**Total**
Video footage (hours)	20	4	10.75	34.75
Macaques (n)	8	3	7	12

All behavioural data with multiple observations were compared and the overall accuracy of the agreements is listed in Table 4.3. Some of the datasets listed overlap, e.g. there are some datasets that both RH and KC observed twice and so both intra- and inter-rater variability has been assessed. Cohen's Kappa is used in observer comparisons, ranging from -1 to 1, with 0 representing no agreement and 1 being perfect agreement. Both intra- and inter-rater variability kappa scores for all observers have a moderate to strong level of agreement (McHugh, 2012). With the exception of inter-rater agreement of behaviour R (86%), intra- and inter-rater agreement in each behavioural category was above 90%. Both intra- and inter-rater variability were highest when predicting P behaviour (98 and 95% respectively).

Table 4.3: Intra- and inter-rater variability of multiple behavioural observations. RH, KC and NG represent different researchers scoring behaviours from video footage.

Observations by	Overall agreement (%)	Cohen's Kappa (κ)	Video footage (hours)	Macaques (n)
RH + KC + NG	72-81	0.63 – 0.75	5	2
RH + KC	88	0.84	15	5
KC + KC	91	0.88	8	3
RH + RH	93	0.90	9.5	4

4.3.1.2 Model validation

The final model was able to successfully predict behavioural classifications from actigraphy data above chance level, with an overall success rate of 69% ($p=2.2 \times 10^{-16}$, 95% CI [67, 70], $\kappa=0.58$). Validation data included 10.75 hours of behaviourally observed data that had not previously been used to build the model. However, 3 of the 7 macaques that formed this dataset had previously been included in model development data.

Table 4.4: Confusion matrix for prediction performance on novel data (10.75 hours from 7 macaques). Each cell represents the accuracy of classifications to each behavioural class. Correct predictions can be found on the diagonal.

		Model prediction				
		IA	**AR**	**GL**	**P**	**R**
Behavioural Observation	**IA**	254	183	3	0	1
	AR	172	849	115	4	2
	GL	0	137	388	83	54
	P	0	4	142	522	16
	R	0	10	24	22	105

Classifications were achieved with reasonable overall accuracy for all 5 behavioural categories (IA= 76%, AR=78%, GL=73%, P=88% and R=79%). Sensitivity values represent the percentage of the time that a behaviour occurs and is correctly predicted by the model. With sensitivity of 60%, 58% and 59% for IA, GL and R respectively, these behaviours are more often misclassified by the model, compared to AR (72%) and P (83%) behaviours. IA and GL behaviours are most frequently misclassified as AR and R often misclassified as GL or P. Of the 7 macaques used for validation, overall accuracy varied between individuals from 52-79%. The 3 macaques with the highest model accuracy (74-79%) spent the most time engaged in pacing behaviour (30-36%).

4.3.2 Case study 1: Pacing

In this case study, the behavioural daily profile of a pair-housed male macaque revealed a pattern of pacing behaviour in response to a particular event. Figure 13 shows a 10-day profile of this macaque, with each day shown as a row and represents the 12 hour light period (7am to 7pm). Every 10 second epoch is represented by a line coloured with the corresponding predicted behaviour. In this instance, actigraphy data were obtained upon request from a researcher who had noted a pattern of pacing behaviour emerging whenever they came to conduct behavioural testing on this macaque. It was challenging for behavioural assessment to be conducted through observation, as the testing regime usually occurred outside of the hours of the support staff and the researcher was preoccupied, testing another macaque at the time of concern.

An extended period of pacing behaviour (green) frequently begins at a similar time every afternoon (~14:00-16:00). A potential cause for this behaviour was revealed when the behavioural testing schedule for this animal (A) and their social partner (B) were considered. In Figure 13, the solid black 'up' arrow indicates when the social partner (B) was removed from their home cage in order to undertake their daily behavioural testing regime. At this time, or shortly after, the pacing behaviour begins. This behaviour continues until macaque (A) is removed from the home cage for their own testing regime, indicated by the black dashed line. The return of (B) to the home cage, dashed black 'down' arrow, does not appear to break the pacing behaviour. This period of pacing behaviour is absent on a non-testing day (12/03/17). On the 08/03/17, there is a longer period between the return of (B) and the start of (A) testing. This was due to a technical fault with the testing equipment. As the two macaques share the same testing set-up, this needed to be fixed before (A) could begin. On this day, a longer period of pacing is observed.

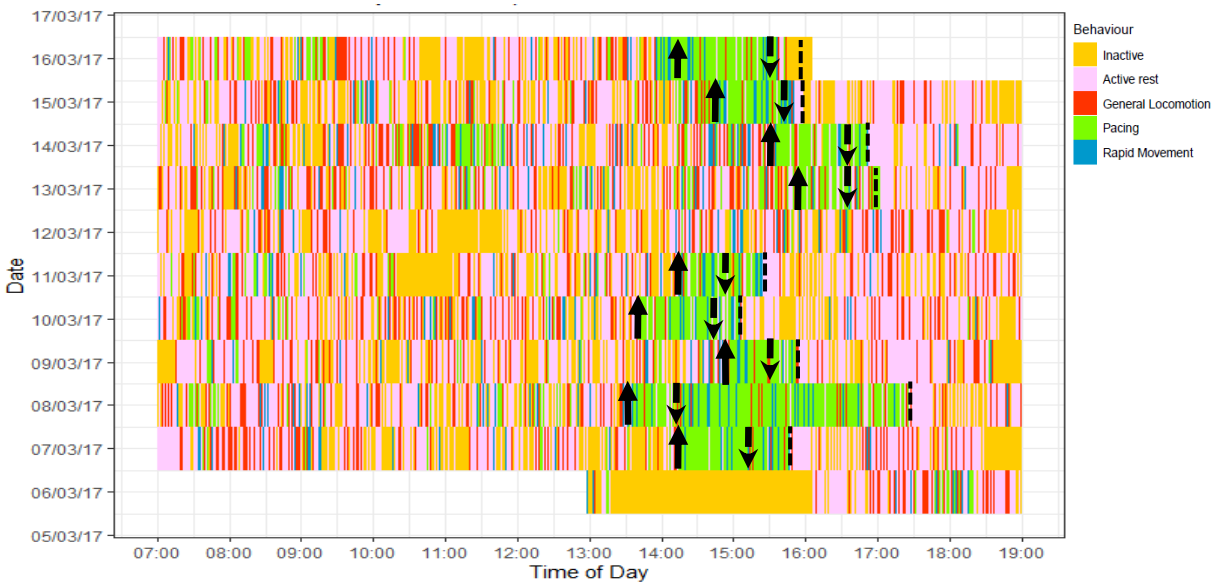

Figure 13: A 10-day profile of a male macaque. Each row is a day (12-hour light period). A behaviour is classified every 10 seconds and is represented by a line in the corresponding colour. Solid black up arrows indicate when the macaque's social partner was removed from the home cage for behavioural testing and the dashed black down arrows represents the partner's return. The black dashed line indicates when this macaque was removed for testing.

4.3.3 Case study 2: Relocation

This case study focuses on the husbandry event of moving a macaque from one holding room in the Oxford facility to another. The accelerometer was deployed for the specific reason of assessing behavioural change following this event. Figure 14 shows the behavioural activity budget of a pair-housed, female macaque whilst housed in room (a) and whilst housed in room (b). Each pie chart (a and b) represents a 5 day (7am-7pm) period, from a Friday to a Tuesday. Chart (b) data were taken 7 days after moving rooms. Before and after the move, the macaque was housed with the same social partner (another female). The experience in each room was the same for the macaque in terms of food given, forage, substrate, cleaning regimes, cage size, environmental factors etc. A difference between the rooms was the number of other macaques housed in them. Room (a) contained only the focal macaque and their social partner but room (b) housed 5 other macaques, 3 of whom were males. Whilst no direct contact was possible with the other macaques in room (b), cage set-ups are such that macaques can see, hear, smell, and indirectly socialise with each other. For example, macaques have often been observed making aggressive facial gestures or lip-smacking (affiliative behaviour) at macaques housed in the cages opposite. The relocation results in an increase in AR and GL type behaviours and a reduction in the time spent IA.

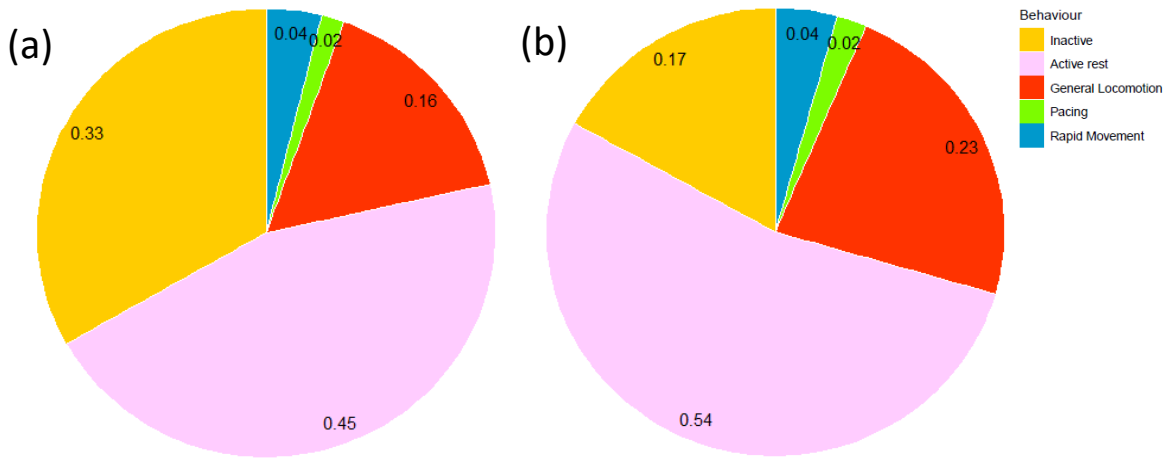

Figure 14: Activity budget of a female macaque before and after a relocation to a new room within the Oxford facility. Pie charts represent 5 days whilst housed in room (a) and 5 days whilst housed in room (b).

4.4 Discussion

4.4.1 Rule-based model

Accelerometers have proven to be an effective tool for obtaining an objective measure of animal behaviour and thus, can be used to build a picture of overall welfare. In this study, actigraphy data have been used to identify 5 species-typical behaviours in macaques used in neuroscience research, with a good overall accuracy of 69%.

Of the 5 behavioural categories, pacing behaviour is predicted with the highest level of accuracy (88%) and sensitivity (83%). As pacing is a common stereotypy in captive macaques (Lutz et al., 2003), the detection and quantification of this behaviour plays a key role in welfare assessment (Mason & Latham, 2004). The presence of pacing does not necessarily indicate compromised welfare. There are multiple causal factors for pacing behaviours; an animal may have a high desire to walk or may pace as a coping mechanism in stressful situations, where pacing actually prevents elevated stress levels (Poirier & Bateson, 2017). However, whatever the causal factor, the presence of pacing does indicate a potentially compromised environment and so is still a useful tool for welfare assessment.

The definition used for pacing in this study may not match that of stereotypical pacing as defined by others. For example, a prior study into the pacing behaviour of research rhesus macaques distinguishes between pacing and agitated locomotion (Poirier et al., 2019). The behaviours are both repetitive patterns of locomotion but differ in their gait and flexibility of the path used. Whilst this distinction can be made through direct behavioural observation, the resolution of the accelerometer data would be unable to detect these differences. Therefore, when pacing is detected by this model, it does not necessarily refer to a stereotypical behaviour.

Other locomotion behaviours (GL and R) were well-predicted by the model. This is consistent with what has been found in other animal studies using accelerometer data, where

locomotor behaviours of differing paces appear to be distinguishable (Fehlmann et al., 2017; Sellers & Crompton, 2004; Watanabe et al., 2005). The behavioural category 'rapid movement' does include some non-locomotor behaviours which could not be distinguished from the locomotor ones. 'Head shake' behaviours involved the macaque rapidly shaking their head from side to side, as if shaking something off. The behaviour was very brief (<1-2 seconds) but due to the neck positioning, the accelerometer is vigorously shook during this behaviour and results in a high actigraphy epoch. This would need to be considered when interpreting R behaviours from macaques.

There is considerable overlap in the activity levels produced by IA and AR behaviours, making them harder to differentiate between. Behaviours from the AR category are mostly linked to the upper limb movements, such as foraging and grooming. The positioning of the accelerometer in a neck collar may mean that the device is unable to detect the magnitude of these movements. Increased behavioural resolution may be achieved by using accelerometers that measures across multiple axis. The actiwatch used here is uniaxial and measures acceleration in the vertical plane only. Some success in identifying similar behaviours have been seen in other primate species using triaxial accelerometers attached to collars (Fehlmann et al., 2017; Sha et al., 2017). Contrastingly, a study using an omnidirectional device was unable to detect arm and head movements in rhesus macaques (Papailiou et al., 2008). Deploying an accelerometer on the arms would most likely provide a better resolution of data from which AR behaviours could be more accurately predicted. Unfortunately, this would not be possible in a NHP species, as easier access to the device could pose a safety risk to the animal.

The accuracy of behavioural predictions using this model varied between individuals. It is common for animals to differ in their behavioural tendencies, within a population or species, and is often referred to as their personality differences (Wolf & Weissing, 2012). As the accuracy of predictions varied with behavioural category, it is likely that the individual variation is linked to

the behavioural differences. In this study, macaques with the highest reported model accuracy also had the highest rate of pacing performance. From observation of this macaque facility, there appear to be macaques that are pacers and those that are non-pacers (I.e. no or rare observation of pacing behaviour). This is also observed in other macaque colonies (Poirier et al., 2019). Therefore, due to high predictive performance of pacing in this model, behavioural assessment of pacer macaques will be more accurate.

There are options available for adjustments that may improve the predictive power of this model. As IA and AR behaviours are often confused, reclassification of the behaviours may make them more distinct. Rather than attempting to capture all instances of IA, focusing on sustained IA instead could reduce the overlap between the two behaviours. This would lose the detection of short inactive periods, which from observation often involve the animals remaining vigilant and engaged with their environment. It should still capture periods of sleep and prolonged rest, which are important behaviours to consider during welfare assessment. Sustained periods of immobility are also associated with the behavioural profile of depressive-like macaques (Camus et al., 2014), so monitoring of these is a key component of welfare assessment.

It may be relatively easy to determine how long or often an animal is engaged in a particular behaviour through automated measures. However, the reason behind it may be different and have very different effects on animal welfare (Rushen et al., 2012). For example, an increase in locomotion behaviours could mean that an animal is exploring more or it could be because they are being chased around or displaced by dominant conspecifics. The resolution of the actigraphy data limits the behaviours that can be predicted and the behavioural categories within this study are broad. For example, AR engagement could mean an animal is foraging, grooming, or picking at a wound, which will have different influences on overall welfare. Providing further context to the data could shed light on which specific behaviour is most likely within a category. In the case of group or pair-housed animals, aligning the actigraphy data of the

individuals could provide some of this context. For example, frequent, brief periods of rapid movement that appear synchronously in a pair may indicate fighting, highlighting a need for closer observation and potential separation. Using the actigraphy predictions alongside existing cage-side assessments of behaviour could also help to provide further context.

Although adjustments could be made to improve model predictions further, the current rule-based system already offers a valuable opportunity to assess the behavioural profiles and welfare of macaques used in research. Examples of how this model can be effectively applied can be seen in the case studies presented.

In addition to the rule-based model, alternative approaches for automated behavioural prediction were trialled. In collaboration with another DPhil student (Sofia Minano Gonzalez) and a research assistant within the lab (Angus Fisk), we explored the option of predicting behaviour from actigraphy data using hidden Markov models (HMM). These have been used previously to analyse accelerometer data in animals (Leos-Barajas et al., 2017). Attempts were made at both an unsupervised and a supervised learning approach. From initial analysis, an unsupervised approach correctly classified behaviours with a success rate of less 50%. This approach was taken no further. The supervised approached proved able to classify activity based on movement intensity, such as slow, medium, and fast-paced movement. However, attempts to classify the species-typical behaviours were unsuccessful. Logistic regression analysis has had previous success with distinguishing cattle behaviours from accelerometers (Yoshitoshi et al., 2013) and was also considered for classifying behaviours in this study. The first developmental stages of this involved prediction of just one behaviour, pacing, from actigraphy data. From this, pacing behaviour was predicted with the similar accuracy (86%) as seen when using the rule-based model. Therefore, the decision was made to focus resources on developing just the rule-based system.

4.4.2 Case study 1: Pacing

In case study 1, a period of pacing behaviour in a male macaque appears to be triggered by the removal of his social partner from the home cage. This could be associated with stress as a result of the social separation (Greco et al., 2017). However, as the pacing behaviour continues following the return of the social partner, this seems unlikely. As behavioural testing of this pair was always conducted consecutively, a more likely explanation is that the social partner's removal triggers an anticipation for this macaque's own testing regime. Anticipation of predictable events, whether positive or aversive, can lead to stress and frustration (Gottlieb, Coleman, et al., 2013). Changes to the order of animal testing or conducting testing in parallel could be a simple refinement that would likely reduce the instance of pacing in this animal. The ability to monitor behaviour continuously and remotely will highlight day-to-day behavioural patterns and could help to identify other similar refinements.

4.4.3 Case study 2: Relocation

In this example, relocation to a room housing more macaques results in an increase in AR and GL type behaviours. The increased socialisation opportunities, although indirect, may be the drive of this change. From personal experience, macaques interacting with their inaccessible roommates (i.e. those they can see but not touch) do so by moving about the front of the cage and engaging in gestures such as head bobbing, threat facing, vocalising, presenting, or lip-smacking. If classifying these into 1 of the 5 behaviours from this study, it is most likely they would fall into either the GL or AR categories. As social companionship is inherently important for the welfare of most primate species (Rennie & Buchanan-Smith, 2006b), this highlights that even the provision of visual, auditory, and vocal interaction opportunities can increase social behaviour. In situations where direct social housing is not always possible, particularly when it would compromise the safety of the animals, it is useful to know that some social needs can still be met.

These are just some examples of how the actigraphy data could be used to explore behavioural changes in response to events that may impact animal welfare. Many other experimental or husbandry related events could also be assessed in this way. For example, it could benefit the post-operative monitoring of pain or detect stress following routine cleaning regimes.

4.4.4 Conclusion

Accelerometers offer an opportunity for animal behaviour to be monitored objectively and continuously, over extended time periods and with minimal human intervention. Although behavioural classes from this model are broad, the behaviours that can be identified are appropriate and important for NHP welfare assessment. This behavioural model has the potential to assess the immediate effects and impact of experimental procedures and husbandry conditions, as well as being able to provide a long-term evaluation of welfare. This system is aimed to be used alongside ongoing cage-side assessments and could provide a substantial contribution to the process of refinement.

Chapter 5 - Cortisol and non-human primate welfare

Quantifying an animals' level of stress may lead to a greater understanding of animal welfare. The stress response activates in reply to a shift in homeostasis, triggered by exposure to certain situations (stressors). This adaptive response is designed to help cope with challenges, such as physiological or psychological disturbance. It is often brief and aims to restore homeostasis (Moberg, 2000).

An essential component of the stress response is the activation of the hypothalamic-pituitary-adrenal (HPA) axis. The HPA axis controls the neuroendocrine response to stressors and causes a cascade of hormones, ultimately leading to glucocorticoid (GC) release into the bloodstream in order to re-establish homeostasis (Matteri, 2000). Thus, detecting GCs and it's metabolites in various biological samples reflects the activity of the HPA axis and can provide insight into both acute and chronic stress.

As well as being a welfare consideration, stress and GC levels affect cognitive performance. Yerkes and Dodson first proposed a model of an inverted U-shaped relationship between performance and arousal (Yerkes & Dodson, 1908). This model suggests that performance in higher-level memory and learning tasks, such as those used in neuroscience studies and relying on the hippocampus and prefrontal cortex, is optimal at moderate stress levels and that very low or very high GC levels associates with poor task performance. This relationship may be influenced by factor such as the duration, timing, and intensity of the stress (Sandi & Pinelo-Nava, 2007). Ongoing assessment of GC levels in animal research models may offer insight into this relationship and could be beneficial to the research outcomes.

Different mediums for cortisol assessment were trialled in this study, to establish which would work best for macaques within a laboratory environment (see section 2.2). As a result of this, hair and faecal samples were collected over several years from rhesus macaques partaking in

neuroscience research. The concentration of cortisol metabolites within these mediums were assessed and will be discussed in terms of their validity as welfare measures and their responses to neuroscience practices.

5.1 Faecal cortisol in rhesus macaques

5.1.1 Introduction

GCs are excreted together with other hormones through faeces in a metabolized form. This provides an opportunity to detect GC metabolites in faeces, which represents an integrated average of plasma free-GC in the time between GC release and excretion (Sheriff et al., 2011). Evidence that faecal GC metabolites correlate with free GC concentrations in the bloodstream and reflect an animal's capacity for stress response suggest that faecal and plasma GCs tell the same story (Sheriff et al., 2010). This has also been reflected in rhesus macaques, where correlation between faecal and plasma cortisol concentrations were observed (Hoffman et al., 2011). Further validation of faecal sampling for GC analysis has been done through the successful detection of physiological manipulations of the HPA axis, such as the ACTH challenge and dexamethasone suppression tests (Palme, 2019).

Faecal samples can simply be collected where they fall, without any need for prior animal training, restraint, or even close proximity to collectors. This makes this method particularly popular in wild animal studies or situations where minimal disruption is essential. There is a time lag between GC release and the ability to detect them in faeces. This is largely a factor of species-specific gut passage rates. In mammals, this lag has been reported as between 20 to 48 hours and in primates, 1 to 3 days (Heistermann et al., 2006; Whitten et al., 1998).

As with other sampling techniques, faecal monitoring of cortisol does not come without its problems and limitations. Within a species, individual differences in faecal cortisol metabolites (FCM) may be observed. This could be a result of multiple factors including age, sex, season, reproductive status, time of sample collection, and defecation rate (Novak et al., 2013). To optimise studies, immediate preservation of faeces after defecation is preferable, ideally by freezing. Whilst this may pose some methodological difficulties, doing so ensures minimal

environmental influence on cortisol concentration, such as bacterial or microbial degradation (Sheriff et al., 2011). Furthermore, faecal GC measures may be ill-suited to assessing the immediate effect of acute stress due to the time lag between stressor and detection of faecal metabolites.

Documenting stress and animal welfare through faecal cortisol analysis is widely used amongst non-human primate (NHP) species. For example, FCM are affected by dominance rank in macaques. This effect appears to vary depending on group stability, with high-ranking males showing elevated cortisol in unstable conditions and subordinates having higher cortisol when groups are more stable (Higham et al., 2013; Ostner et al., 2008). Additionally, elevation in FCMs has been recorded in response to capture stress in gray mouse lemurs (Hamalainen et al., 2014) and following immigration into a new group in crested macaques (Marty et al., 2017). Faecal GCs have also been useful for quantifying the stress response to experimental procedures experienced by rhesus macaques used in research, such as jacketing and quarantine housing (Field et al., 2015; Jackson et al., 2023).

As discussed in an earlier chapter, 'Chapter 3 - The impact of general anaesthesia on non-human primates,' GA has an essential role to play in neuroscience research using NHPs. However, it does also have an impact on animal welfare. Anaesthetic stress is well-documented and GA has even been used as a controlled stressor to evaluate the stress responsiveness of faecal cortisol (Whitten et al., 1998). The stress response induced by a GA is believed to be associated with the disorientation experienced upon entering and exiting the GA, rather than a direct influence on the HPA axis. The impact of GA alone is rarely documented, as often it is accompanied by additional stress, such as a surgical procedure. Nevertheless, there is some evidence of elevated GC levels following anaesthesia alone (e.g. in mice (Hohlbaum et al., 2018) and chimpanzees (Whitten et al., 1998)). Furthermore, faecal GC metabolites peak following an anaesthetic to facilitate a physical examination in a cheetah and minor (laparoscopic) surgical procedures in a

clouded leopard (Young et al., 2004). Comparative studies in mice revealed elevated serum GC concentrations following both an anaesthesia alone and an anaesthesia in addition to a minor surgical procedure (vasectomy). However, these levels remained high for a longer period following surgery compared to just a GA (Jacobsen et al., 2012a). An earlier study did not detect these differences between the two procedures when assaying faecal GCs instead (Jacobsen et al., 2012b).

Ascertaining the stress response to common neuroscientific procedures performed on rhesus macaques allows for better understanding of welfare impact and highlights opportunities for refinements. Macaques followed on this study were repeatedly exposed to GA, both for the purpose of conducting an MRI or a surgical procedure. Assessment of faecal cortisol metabolites (FCM) provides a non-invasive, simply attainable measure to understand the magnitude and breadth of the stress response in these situations.

In this study, I aim to determine whether changes in FCM are detectable following administration of GA, providing insight into the acute stress response elicited by this stressor. To investigate this, FCM are observed before and after two different GA events. An MRI event involved GA alone, with no tissue trauma experienced by the macaques. A surgery event involved GA and tissue trauma, performed to affix a headpost to the skull of the macaque which was required for restraint in later neuroscience protocols (see section 2.4 for further details). In addition, the ranges of FCM in macaques housed in the Oxford neuroscience facility is discussed. These FCM ranges are inclusive of stressful events, such as experimental procedures and husbandry related stressors (e.g. disruption to social housing, relocation to a new room).

5.1.2 Methods

For collection and assaying methods, see section 2.2.3.1. Assessment of cortisol concentration in faeces was done through detection of a cortisol metabolite, 11ß-hydroxyetiocholanolone.

Surgical interventions assessed in this section were restricted to headpost (HP) surgeries only. Section 2.4.6 and 2.4.7 of this thesis describes the MRI and surgery procedures in detail.

Total subjects in this study were all adult male rhesus macaques. All macaques were housed at the Oxford neuroscience facility and experienced similar, but not always identical, experimental protocols. Samples were collected opportunistically, and general ranges stated are inclusive of all data assayed. Faecal samples collected to form this dataset could be influenced by any of the protocols stated in the 'Neuroscience methods' section (2.4) of this thesis.

To assess the effect of GA on FCM, I used any faecal sample that had been assayed and was within 7 days (before or after) of a direct GA event.

All macaques have extensive individual history files available, detailing key information and monitoring the performance of regulated procedures. These files are a requirement under the A(SP)A (Animals (Scientific Procedures) Act, 1986) for specially protected species, such as NHPs. These records were the source for information pertaining to the individual animals, such as their age, dates of general anaesthetics etc.

5.1.2.1 Statistical analysis

All statistical analysis and figures were created in R (version 4.1.0) with RStudio (R Core Team, 2021). Plots were created using *ggplot2* (Wickham, 2016).

The *friedman.test* function from the *stats* package (R Core Team, 2021) was used to determine statistically significance differences in FCM, following MRI or surgical events, across 3 time periods (baseline, post-days 2&3 and 4&5). Post-hoc analysis was performed using the *pairwise.wilcox.test* function, also from the *stats* package, and Bonferroni correction was applied.

Baseline FCM was determined by taking the mean of all the available data from 7 days prior to the GA event. Some exploratory data analysis showed an increase in FCM on at least days 2 and 3 immediately following the GA intervention. No change in FCM was observed on days 0 and

1. This was expected due to the temporal lag between the circulation of cortisol in the blood and its appearance in the faeces. Therefore, further analysis focused on the change from baseline FCM to that observed on day 2 and beyond.

5.1.3 Results

Range and mean of FCM from this study can be found in Table 5.1. All data are represented in the stated ranges and averages, regardless of whether the faecal sample is associated with a stressor, e.g. a surgery, or not. To illustrate results in Figure 16, 5 extreme values were removed to improve the plot. The extreme FCMs were as follows: 14278, 13283, 10629, 8647 and 8269 ng/g from macaques PE, PE, VS, TM and TV respectively. These values were included in the rest of the analysis in this study. Results are separated by individual in Figure 16. Many animals show broad FCM ranges and some animals appear to display overall higher FCM values (VL, WD, WG), suggesting large amounts of variation both within and between individuals.

Table 5.1: Descriptive statistics of the datasets used to analyse cortisol metabolites in faeces. Data are represented as the number of events or macaques (n) or as means ± standard deviation, followed by the range [in square brackets].

FCM ranges	Number of macaques (n)	36
	Age ($\bar{x}$ years ± SD; [range])	6.0 ± 1.6 [2.6-10.4]
	FCM (ng/g $\bar{x}$ ± SD; [range])	1157.5 ± 2125.7 [42-14278]
MRI events	Number of MRIs (n events; [n individuals])	42 [12]
	Length of anaesthesia ($\bar{x}$ hours ± SD; [range])	4.9 ± 1.9 [1.0-9.2]
	Age ($\bar{x}$ years ± SD; [range])	6.1 ± 1.5 [3.5-10.2]
Headpost events	Number of surgeries (n events; [n individuals])	18 [18]
	Length of anaesthesia ($\bar{x}$ hours ± SD; [range])	5.3 ± 0.9 [3.2-6.4]
	Age ($\bar{x}$ years ± SD; [range])	6.2 ± 1.0 [4.5-7.8]

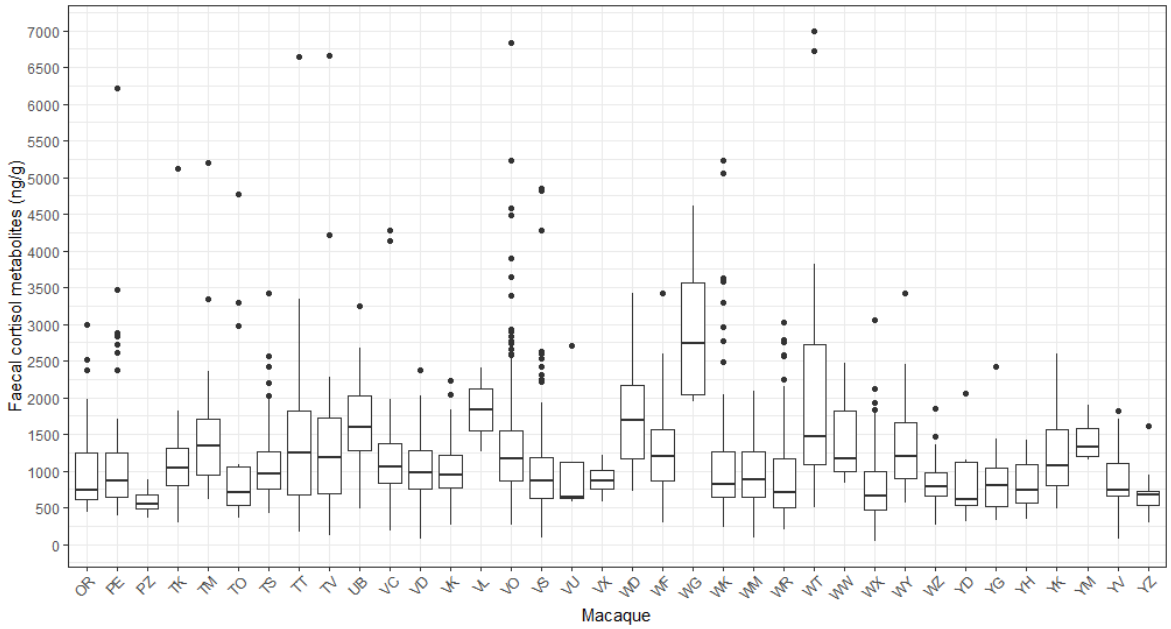

Figure 15: Faecal cortisol metabolites of male macaques. All cortisol values are included, regardless of proximity to a stressful event. 5 extreme values (>7500) have been removed for better visualisation. Black dots represent outliers.

A longitudinal display of the FCM of two macaques is seen in Figure 15. These macaques had the longest period of FCM analysed (4 years) and show a visual representation of FCM over time. During these 4 years, samples were often collected in response to a known stressor and the data represented in Figure 15 is a mixture of baseline and response values.

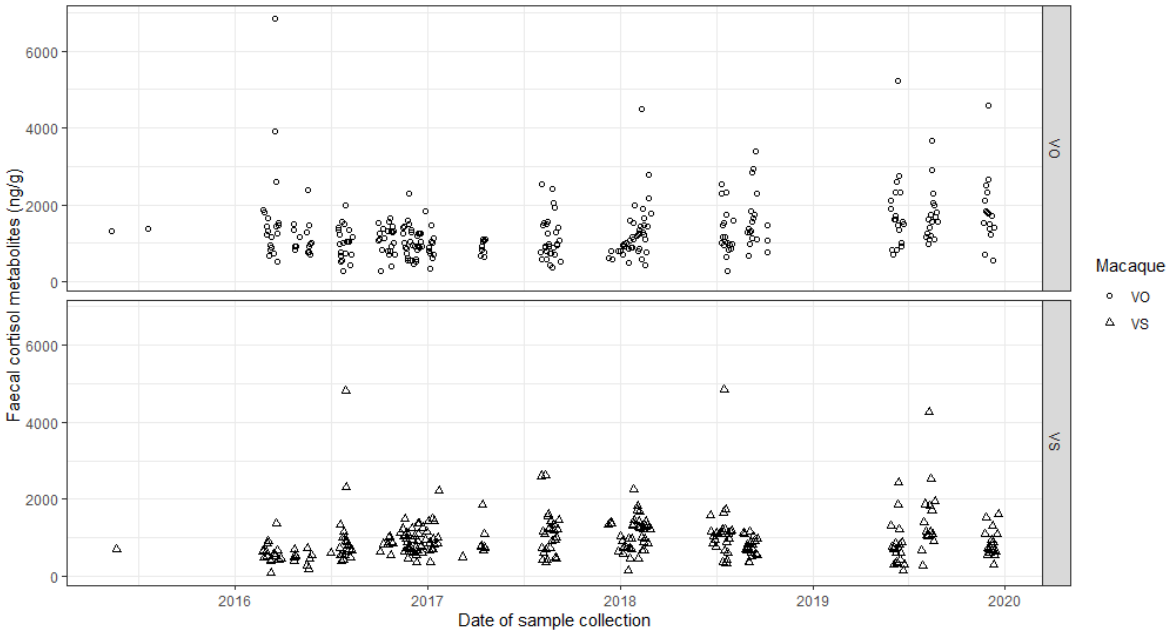

Figure 16: Faecal cortisol metabolites of two male macaques (VO and VS) over time. One extreme value (>10000) has been removed for better visualisation.

Figure 17 shows the FCM response to MRI procedures under GA. Descriptive statistics for the data can be found in Table 5.1. Baseline values represent the mean FCM of individual macaques from up to 7 days pre-MRI intervention, a total of 203 cortisol results contribute to this estimate, inclusive of day 0. Post-Days 2&3 and 4&5 represent either day 2 or 3 (8 interventions) and day 4 or 5 (11 interventions) post-MRI respectively or a mean cortisol level over the 2-day periods where 2 samples were available (34 or 31 interventions respectively). Friedman's rank sum test, comparing cortisol level by pre or post intervention stage whilst considering individuality between animals, implied some significant difference when comparing all 3 parameters (Friedman chi-squared = 20.333, df=2, p=0.000038). Post-hoc analysis using Pairwise Wilcoxon signed rank test (Bonferroni corrected) revealed a significant change in FCM on days 2 and 3 post-MRI compared to baseline levels (1.72-fold increase, p=0.00014 for post-days 2&3). No difference in FCM was observed between baseline and days 4 and 5 post-MRI (p=0.268 for post-days 4&5). Interestingly, a significant difference between both post-MRI periods was observed (p=0.00000026).

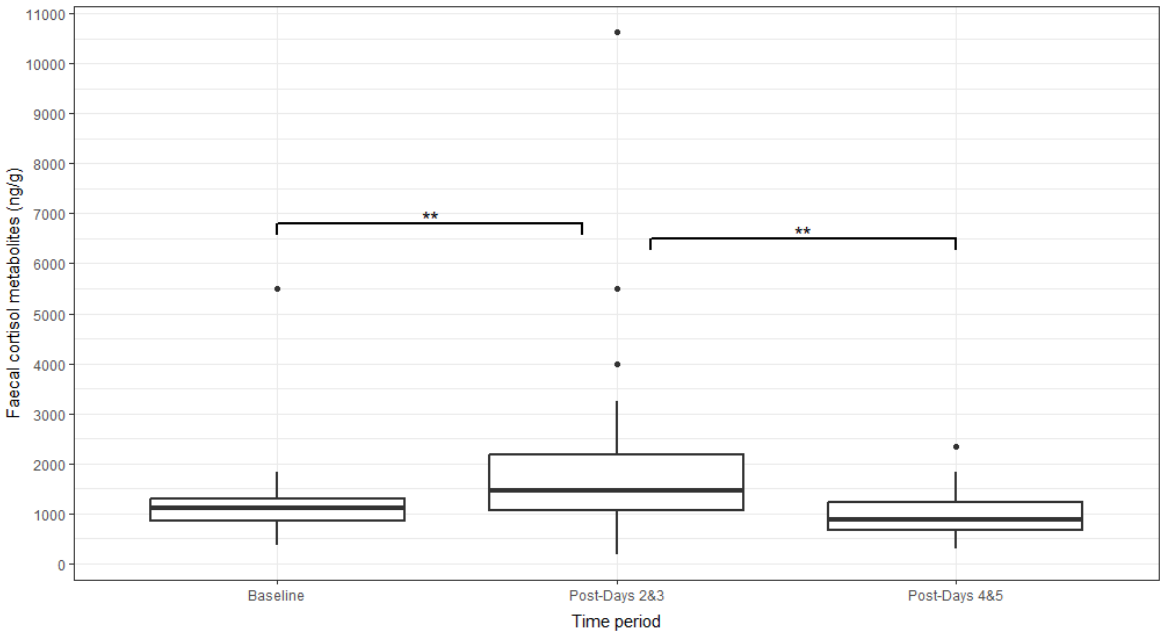

*Figure 17: Faecal cortisol metabolites in response to a general anaesthetic administered for an MRI procedure (n, individuals=12, events=42). MRI was performed on day 0. 'Baseline' comes from data collected up to 7 days prior to the event (inclusive of day 0). 'Post-Days' is the FCM from days 2 and 3 then 4 and 5. ** indicates p<0.001 (pairwise Wilcoxon signed rank rest). Black dots represent outliers.*

Figure 18 shows the FCM response to headpost surgeries. Descriptive statistics for the data can be found in Table 5.1. Baseline values represent the mean FCM of individual macaques from up to 7 days pre-surgical intervention, a total of 79 cortisol results contribute to this estimate, inclusive of day 0. Post-Days 2&3 and 4&5 represent either day 2 or 3 (4 interventions) and day 4 or 5 (2 interventions) post-surgery respectively or a mean cortisol level over the 2-day periods where 2 samples were available (14 or 16 interventions respectively). Friedman's rank sum test, comparing cortisol level by pre or post intervention stage whilst considering individuality between animals, implied some significant difference when comparing all 3 parameters (Friedman chi-squared = 18.1, df=2, p=0.00012). Post-hoc analysis using Pairwise Wilcoxon signed rank test (Bonferroni corrected) revealed significant increase in daily FCM following headpost surgery compared to baseline levels on post-days 2&3 (2.83-fold increase, p=0.00023) and post-days 4&5 (1.9-fold increase, p=0.00011). Furthermore, a significant difference in FCM was found when comparing the two post-intervention parameters (p=0.048).

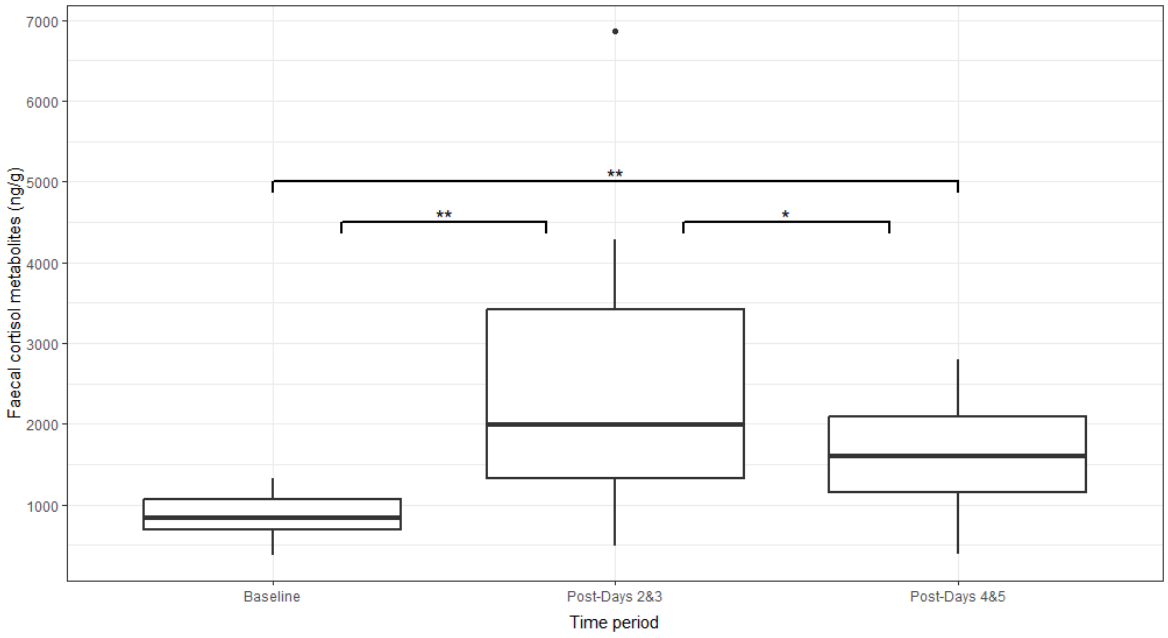

*Figure 18: Faecal cortisol metabolites in response to a general anaesthetic administered for a headpost surgery (n=18). Surgery was performed on day 0. 'Baseline' comes from data collected up to 7 days prior to the event (inclusive of day 0). 'Post-Days' is the FCM from days 2 and 3 then 4 and 5. * indicates p<0.05 and ** indicates p<0.001 (pairwise Wilcoxon signed rank rest). Black dot represents an outlier.*

5.1.4 Discussion

FCM is a commonly used measure to assess the stress response to a variety of different situations, with the aim of gaining a better understanding of overall animal welfare. In this study, faecal samples were collected opportunistically and longitudinally, following the lives and significant events experienced by rhesus macaques undertaking neuroscience research. From the data obtained, I was able to determine the range of macaque FCM within the facility and explore the acute stress response to GA events.

5.1.4.1 FCM in research macaques

The FCM values observed in this analysis span a wide range and have a large standard deviation, suggesting that FCM measures were very varied in these samples. The quoted ranges are inclusive of samples collected in close proximity to experimental procedures and husbandry events, such as social disruption. Thus, ranges consist of both baseline FCM measures and some reflections of FCM responses to these stressors.

Comparisons between the ranges found in rhesus macaques in this study and that of other studies cannot be made. This is because study assay methodologies should be identical, i.e. conducted in the same laboratory using the same methods in all aspects (Michael Heistermann, personal communication, November 2019). To my knowledge, the laboratory used for assaying our samples has not conducted any other analysis on captive-housed rhesus macaques. However, they have reported ranges of free-ranging rhesus macaques. In these, they report FCM levels of 150-800 ng/g in males (Higham et al., 2013) and 600-1000 ng/g in females during the first month of gestation (Brent et al., 2010). These ranges do overlap with the ranges observed in this study but FCM had a very broad range (42-14278 ng/g, mean ± SD = 1157.5 ± 2125.7 ng/g).

FCM in animals may fluctuate and show significant increases during baseline conditions, as there are so many factors that can induce a stress response. FCM variation across days with no known cause is observed in many species (Michael Heistermann, personal communication,

November, 2019). Therefore, it is not unusual that such a vast range of FCM values is observed in this study, even within the 'baseline period' prior to MRI and surgical events. Although specifically taken to avoid the influence of major experimental stressors, these baseline periods are still effected by stressors associated with 'normal' laboratory life. This may include regularly occurring events that impose stress or discomfort, such as daily testing regimes which involve social separation and restraint (Pfefferle et al., 2018). Stressors linked to animal husbandry, such as a fight between conspecifics, could also not be controlled for and occurrence of such during the baseline period may influence FCM values.

Although the data from this study were collected opportunistically, due to the controlled environment in which these macaques live, some factors could be easily standardised. This study displayed consistency in the context of sex and diet of macaques. As well as in the time of sample collection and delay between collection and preservation, all of which could influence FCM (Novak et al., 2013). Therefore, the vast range of FCM observed is unlikely to be a result of these factors. However, elements outside of the control of this study may have contributed to the observed levels of variation. Social stress and dominance status may influence faecal cortisol (Higham et al., 2013; Ostner et al., 2008) and there may be difference reflected in animal age (Rauch et al., 2014). Furthermore, samples in this study were stored at -20°C for up to 4 years before assaying. The effects of storage times of this length may not be available but the stability of cortisol metabolites in frozen faeces has been validated for up to two years (Hunt & Wasser, 2003).

5.1.4.2 FCM and general anaesthesia

In this study, an increase in FCM was observed in response to a GA, which was administered either for an MRI or a surgical procedure. This quantifies the stress response to a GA and highlights the ability of FCM to detect an acute stress response. Immediate cortisol response was assessed in children using serum and saliva sampling techniques and elevated cortisol levels

were observed during emergence and recovery (~70 minutes post-emergence) from a GA (Rains et al., 2009). Measures of faecal cortisol are less susceptible to brief fluctuations in GCs, and so it was unclear whether these acute responses would be detectable in faecal samples. Some previous studies have reported observing the stress response to a GA in faecal GC measures (Hohlbaum et al., 2018; Whitten et al., 1998), where others have been unable to detect it (Jacobsen et al., 2012b).

Following both MRI and HP surgeries, there was a statistically significant increase in FCM on days 2 and 3 post-GA, compared to baseline. This change to FCM was larger following a surgical event compared to an MRI event. Furthermore, there is a significant increase from baseline in cortisol levels on days 4 and 5 post-surgery but no such effect was seen post-MRI. This result implies that surgical intervention has a more profound impact on macaque stress than the administration of GA alone, when administered for the purpose of an MRI. A similar finding in mice revealed an elevated faecal corticosterone response following a vasectomy surgery but not following a GA alone (Wright-Williams et al., 2007). This indicates that the surgery itself was the primary driver of the extended endocrine response. These results are expected as surgical interventions include additional physiological stressors, such as tissue injury and post-operative pain (Desborough, 2000). Furthermore, according to A(SP)A guidelines, GA for MRI only is a brief and transient experience, categorised as a 'mild' procedure, as long as there are no additional adverse effects (for example mobility issue during recovery; trauma to placement of stereotactic headholder). Surgery is considered to have a higher degree of PSDLH, and is typically classed as 'moderate' severity classification (Home Office, 2014b).

In addition to the surgical procedure itself, post-operative recovery from a surgery can impose further detriment to the welfare of macaques. A series of stressful components may accompany surgical recovery. These include the administration of analgesic and antibiotic drugs, habituation to the headpost per se and temporary separation of socially housed individuals

(Drucker et al., 2015). These additional stressors may create a cumulative effect to result in a prolonged elevation of FCM, lasting at least 5 days post-op. Following an MRI event, the macaque would not typically experience these recovery stressors, and this is most likely why FCM elevation is not observed on days 4 and 5 post-MRI. In this study, we do not see a return to baseline FCM levels following a headpost surgery. However, a significant reduction in FCM between days 2 and 3 post-op and days 4 and 5 post-op implies HPA axis activity is returning to pre-intervention levels. To confirm this, further investigation over a longer post-operative period is required.

Following on 4 and 5 days from an MRI procedure, we observed a decline in FCM. Whilst not statistically significant, the FCM observed on those days look like they may have even dropped below the baseline levels. As with the surgical procedures, this could be influenced by the recovery period post-GA. Unlike surgery, however, MRI recovery would usually not invoke further stressors but may eliminate some of the 'normal' day-to-day stressors' macaques experience. Usually, following a MRI event, macaques would be assigned a couple of days of rest, where no neuroscience testing protocols took place. This could eliminate the potential stress caused by the restraint, temporary social separation, and cognitive challenge that these testing regimes involve. Social separation within the home cage could also be a contributor to the differences between FCMs on days 4 and 5 between GA procedures. If these GA events are performed on socially housed animals, they are often kept separate until animal care staff have deemed them to be well enough to regroup. With macaques recovering from MRI, regrouping tends to happen much sooner. For example, following an MRI, macaques are often re-grouped several hours (<1 day) after returning to the home cage but following surgery it is more likely to be 1-2 days. This would reduce any stress caused by social separation but also could offer a social buffering opportunity, providing a companion to assist with recovery (Novak et al., 2013).

Due to the opportunistic nature of data collection some confounds of the data should be acknowledged for clarification. Within the MRI dataset, there are two animals (observed in Figure

15) that experienced more than half of the MRI events recorded between them (each animal underwent 12 MRI events). Although the effect of repeated stressors has not been tested in this study, it is possible that subsequent exposures may lead to either an adaption or sensitisation to that stressor (von der Ohe & Servheen, 2002). The type of response experienced may vary and could be influenced by numerous factors. For example, repetition of consistent stressors, occurring at regular intervals, are likely to be more easily adaptable than severe and irregularly timed stressors. Therefore, with these two animals comprising such a large amount of the MRI dataset, not only could any effect of individual variation skew the results but any influence of repetition on the HPA axis may also drive the data. Although the sample size may be too small, further exploration into the data of these two animals could provide insight into the cumulative effect of frequent GA exposure. Finally, the length of the MRI events in this study varied from 1 to 9 hours. As observed in macaque activity levels in a previous chapter, 'Chapter 3 -The impact of general anaesthesia on non-human primates' the length of a GA may influence the recovery period. Duration of a GA is a risk factor for poor quality recovery and prolonged emergence times (Loomes & Louro, 2021; Sinclair & Faleiro, 2006). Further investigation, to establish whether FCM varies with the length of anaesthesia, could add value to this dataset and the overall understanding of the welfare impact to macaques.

5.1.4.3 Further study

The outcome from this study highlights the benefits of faecal sample collection as a tool to assist in the welfare assessment of NHPs in a captive environment. Whilst using FCM does not give a full picture of welfare, its simple methodology can be easily implemented alongside neuroscience studies and could contribute to the understanding of stress experience.

FCM analysis could be employed to quantify the stress response to other common neuroscience practices. This could include exploring other surgical procedures, of varying severity, the influence of different testing regimes and the effects of husbandry procedures. For

example, some research in macaques observed changes in FCM associated with social factors, such as dominance rank and social stability (Higham et al., 2013). Furthermore, there is great emphasis on the avoidance of single housing in research macaques as social deprivation can be severely detrimental to psychological health. Whilst this effect was not considered in this study, it could be possible that FCMs are influenced by social housing situation. Previous studies in rhesus macaques have detected both an increase (Lilly et al., 1999) and no affect in plasma cortisol concentrations during single-housing (Schapiro et al., 1993), as well as no affect in faecal cortisol (Jackson et al., 2023). However, singly housed animals showed increased incidence of stress-associated behaviours, such as stereotyped and self-injurious behaviours (Lutz et al., 2003). These effects suggests that the standard functioning of the HPA axis could be compromised, effecting baseline cortisol levels. Furthermore, socially housed animals may benefit from social buffering, where the presence conspecifics improves recovery from distress (Kikusui et al., 2006). Using a measure such as FCM can offer the opportunity to assess these social influences without interference. Macaques housed in a research environment will become naturally exposed to social stress and although efforts to avoid it are maximised, single housing is often still required.

As neuroscience research often lasts several years, it is essential to consider the cumulative experience of macaques throughout their lifetime (Pickard, 2013). Faecal sampling offers a simple methodology and stable, long-term storage options which make it ideal for regular, longitudinal data collection.

5.1.4.4 Conclusions

Assessing the stress using faecal samples has proved a valuable tool for offering insight into the welfare of rhesus macaques enrolled in neuroscience research programs. Exploration of one experimental protocol, revealed an acute stress response to the administration of GA. FCM were elevated when GA was given for an MRI or a surgical procedure. This response was observed 2 days after the GA event, aligning with the expected delay caused by the time taken to excrete

cortisol metabolites in this species. Returning to baseline cortisol levels occurred after 3 days following the less invasive, MRI events but didn't fully return to baseline 5 days after the headpost surgery. This suggests that the recovery from a surgical intervention imposes more stress on macaques than recovery from a GA alone.

Employment of this methodology offers researchers a simple tool that involves minimal disruption to their research protocols and requires little time investment. However, it does offer the opportunity to quantitatively assess the stress response to many neuroscientific protocols and, used alongside other welfare parameters, could offer insight into the welfare state of research animals. Understanding the welfare impact of specific scientific protocols offers opportunity to implement refinements that may improve the lives of animals.

5.2 Hair cortisol in rhesus macaques

5.2.1 Introduction

Obtaining a long-term endocrine record can provide key insight into the development of, and contributors to, an animal's chronic stress. Chronic stress has proven to be a high-risk factor leading to the deterioration of health in humans and animals. When stress exceeds an animal's ability to cope, it puts them at increased risk of many physical and mental illnesses, such as cardiovascular disease and psychiatric disorders (Barnett & Hemsworth, 1990; Hurst et al., 1976; Salleh, 2008). Whilst cortisol measures from urine, saliva, and faeces have a part to play in stress assessment, they typically focus on cortisol changes within a narrow time frame. To assess the level of stress long-term, repeated sampling of individuals would be required, which brings with it many logistical difficulties. Accumulation of hormones in the hair follicles as they grow offers a timeline of cortisol secretion and therefore an assessment of HPA axis activity, over the lifetime of the hair (Carlitz et al., 2014; Davenport et al., 2006). Experimentally induced cortisol release, via the administration of adrenocorticotropic hormone (ACTH), has been conducted in multiple species and validates the medium of hair as a representative for the assessment of the long-term function of the HPA axis (Heimburge et al., 2019). Therefore, the collection of hair samples could be a valuable tool to monitor the chronic stress that might be associated with exposure to the multiple experimental procedures experienced by research animals.

Cumulative experience refers to the consideration of the net impact of all events, both positive and negative, that a research animal experiences over their lifetime (Animals in Science Committee, 2017). Measurement of hair cortisol concentration (HCC) offers a medium to observe and quantify repeated exposure to particular experiences, and determine whether regular stressors have a long-term impact on HPA axis activity. As cortisol accumulates through the lifetime of the hair, a single HCC measure provides information about the workings of the HPA axis over several weeks. Although hair sample collection is brief and non-invasive, there may still

be acutely stressful elements involved, such as the use of restraint for the purpose of neuroscience training and/or data collection. However, HCC provides an average of cortisol exposure over the hair growth period (Sheriff et al., 2011). Thus, brief and infrequent acute stressors will not confound HCC result nor will the circadian rhythmicity of cortisol that is observed in other sample mediums (Stalder et al., 2012).

Further benefits of measuring HCC included its stability over time. Samples can be stored at room temperature for extended periods of time (Sheriff et al., 2011). Although relatively stable when stored, there is some evidence of segmental decline throughout the hair strand. This is thought to be caused by exposure to external stimuli whilst the hair is still growing, such as UV radiation, hair washing, and hair treatment (e.g. bleaching, cosmetics) (Meyer & Novak, 2012). As this study uses laboratory housed macaques, who are permanently indoors, their exposure to such elements is significantly reduced and thus HCC is unlikely to be strongly affected. Furthermore, cortisol concentration in hair samples has been validated in rhesus macaques used in research; showing significant correlation to salivary cortisol and sensitivity to major, prolonged stressors (Davenport et al., 2006).

Some of the key challenges facing a laboratory housed NHP include social constraints, reduced living space, use of restraint, and administration of general anaesthesia (GA). GA disrupts many physiological processes and induces physical stress. Previous studies have looked at the acute effect of GA on welfare, with some studies using cortisol to highlight acute stress following a GA (plasma and saliva cortisol following non-surgical GA (Rains et al., 2009), urinary cortisol in chimpanzees (Anestis, 2009)). However, little is known about the long-term impact of GA on animal welfare. Care needs to be taken when assessing the impact of GA as in most cases GA is used for the purpose of conducting a surgical procedure. Therefore, it would not be possible to determine whether stress is induced by the GA or by recovery and surgical stress to which GC increase is an expected endocrine response (Cusack & Buggy, 2020).

In addition to their experimental protocols, challenges of living within a laboratory environment may also induce stress, acute or chronic, in NHPs. Macaques are social species and the absence of sufficient social contact is considered to be a significant psychological stressor (Rennie & Buchanan-Smith, 2006b). It is a legislative requirement for NHPs used in research to be socially housed with conspecifics, whenever possible (Animals (Scientific Procedures) Act, 1986). However, unstable social groups may also be detrimental to NHP welfare, increasing levels of social stress as a result of an inability to avoid conspecifics in the restrictive living quarters of a laboratory setting. From the literature, there are many reports of higher HCC levels associated with social stress. In primates, this includes elevated HCC upon receipt of social aggression in great apes (Carlitz et al., 2014; Yamanashi et al., 2013) and in response to social instability and population density in rhesus macaques (Dettmer et al., 2014; Vandeleest et al., 2019).

Whilst many previous studies report hypercortisolemia as a response to chronic stress, there is some evidence of HPA axis activity suppression, leading to hypocortisolemia. Studies of humans suffering from Post-traumatic stress disorder (PTSD) have often reported lower than normal cortisol levels (Fries et al., 2005). Social stress as a result of unstable social groupings in rhesus macaques lead to reduced basal plasma cortisol levels and reduced cortisol response to acute stressors (Capitanio et al., 1998).

In this part of the study, I aimed to assess the utility of hair cortisol as an indicator of chronic stress in macaques used in neuroscience research and to investigate questions associated with animal welfare. First, I asked whether HCC measures were stable over time (storage effect). I then investigated whether periods of known stressors produced peaks of HCC independent of storage time. Thereafter, I examined questions associated to animal welfare, e.g. whether the length of time in the research facility, surgical procedures, and being singly housed affected HCC. Two individuals were treated with varying concentrations of prednisolone (a steroid drug

treatment) during hair sample collection. We therefore also took the opportunity to investigate the effect of this treatment on HCC.

5.2.2 Methods

For collection and assaying methods, see section 2.2.4.1.

Subjects in this study were all adult male macaques, aged 5.5 to 12 years of age. Due to the method for sample collection, macaques could not be enrolled on this study until they had undergone a surgical procedure to attach a head fixation device (see section 2.4.7). Therefore, all animals had been living in the neuroscience facility at Oxford for at least 1.5 years before sample collection began (see Table 5.2 for further descriptive statistics).

All macaques have extensive life records available, detailing key information and monitoring the performance of regulated procedures as required under A(SP)A 1986. These records were the source of information pertaining to the animals.

Unrelated to the study, during the course of sample collection, two macaques (PC, PZ) displayed procedure-unrelated clinical symptoms of atopia (areas of red, irritated skin) requiring treatment with corticosteroids (prednisolone). This treatment was administered as standard, administration of oral tablets in a palatable 'treat' which was consumed voluntarily. Prednisolone is a synthetic glucocorticoid and one of its side effects is listed as being 'HPA axis suppression' (Papich, 2021). Due to this, any HCC from these two animals have been removed from the data before creating descriptive statistics, plots, and models. The two macaques were treated with prednisolone for 23 and 20 months. Frequency of administration and dose was adjusted based on the individual animal's condition as per veterinary advice. Of that treatment time, 13 months overlapped hair sample collection. In total, 19 hair sample results (11 from one animal and 8 from the other) were affected by the treatment, but the decision was made to remove all data from these animals as although treatment had been discontinued, the underlying causes requiring

treatment could have continued to alter cortisol levels (chronic inflammatory response). Hair cortisol measures from these two macaques have been plotted separately against dosage of prednisolone received to observe the effect of treatment on HCC.

5.2.2.1 Statistical analysis

Hair cortisol results from EIA assaying from 8 male macaques were not analysed in depth but basic statistics are quoted. Cortisol results from LC-MS/MS assaying were analysed for 24 male macaques. All statistical analysis and figures were created using R (version 4.1.0) with RStudio (R Core Team, 2021). Plots were created using the package *ggplot2 (*Wickham*,* 2016*).* Hair cortisol measures were not normally distributed but measures approached normality after log transformation.

To test the influence of various intrinsic and extrinsic factors on HCC, linear-mixed models (LMM) were employed. LMMs were generated using the *lmer* function from package *lme4 (*Bates et al., 2015*),* and model outputs were extracted using *summ* function from the *jtools* package (Long, 2020). Data were tested for outliers using *outlierTest* from the *car* package (Fox & Weisberg, 2019). Model residuals appeared normally distributed upon observation of qqplots.

To investigate if the length of storage affected the level of HCC, I visually inspected HCC profiles from all macaques over time and conducted a LMM with 'length of storage' as a fixed factor and individual ID (Identity) and social group ID (Social group) as random factor.

A further LMM was employed, including the age of the animal when they arrived at the Oxford facility (Age on arrival), the length of time they had been housed at the Oxford facility (Time in Oxford) and whether they were singly housed (binary: 1 for single housing, 0 for group or pair housed). The LMM further included 'MRI' or 'Surgery' as fixed effects, which pertains to whether the macaque received a general anaesthetic (GA) within the period of hair sample collection (binary: 1 or 0/yes or no). Individual animal ID (Identity) and ID of their social group (Social group) were included as random effects. Due to collinearity, the 'length of storage' was

not included in this model and instead run separately. There were some occasions where sample collections occurred very close to GA interventions. As cortisol is incorporated into the hair before the hair breaks the skins surface, there is a time delay until the hair sample containing cortisol pertaining to those events can be collected. Therefore, association with the hair sample was adjusted such that if an event occurred 7 days or less prior to a hair sample collection, the event was not associated with that hair sample but with the next sample collected.

5.2.3 Results

Hair cortisol concentrations (HCC) from samples analysed using the EIA assaying method ranged from 9.1 to 130.7 pg/mg hair (mean: 37.9 ± 15.9 (SD)). Results from this assay are not illustrated in any of the figures in this section.

HCC from LC/MS analysis across 22 macaques (animals having prednisolone treatment, PZ and PC, were removed) averaged 19.3 pg/mg (see Table 5.2 for more details). Descriptive statistics of dataset parameters can be found in Table 5.2. Results are illustrated per individual in Figure 19. Inter-individual differences become apparent, with measures from PG, PR and VM being lower than from the other animals. One extreme outlier from macaque OL, 171.72 pg/mg, was removed from the dataset.

Table 5.2: Descriptive statistics of the datasets used to analyse hair cortisol concentration using LC/MS assay. Data was represented as the means ± standard deviation, followed by the range [in square brackets], for data derived from either the macaques sampled or the hair samples themselves. Data pertaining to the events that occurred during sample collection is presented as the number of events (n events) and the number of different macaques experiencing those events (n individuals). 'Number of surgeries' consists of several different neuroscience procedures. These are listed at the end of the table and further details on the surgical procedures can be found in section 2.4.7 of this thesis.

Macaques	Age ($\bar{x}$ years ± SD; [range])	8.7 ± 1.5 [5.5-12]
	Age on arrival in Oxford ($\bar{x}$ years ±SD; [range])	4.2 ± 0.9 [2.4-5.6]
	Time since arrival in Oxford ($\bar{x}$ years ± SD; [range])	4.5 ± 1.8 [1.6-9.6]
	Duration of sample collection ($\bar{x}$ years ±SD; [range])	2.5 ± 1.0 [0.6-3.9]
Hair samples	Length of sample storage ($\bar{x}$ years ± SD; [range])	3.1 ± 1.0 [1.2-5.2]
	Cortisol concentration (pg/mg hair $\bar{x}$ ±SD; [range])	19.3 ± 8.5 [3.2-94.5]
Events	Number of MRIs (n events; [n individuals])	37 [13]
	Number of health screens (n events; [n individuals])	9 [8]
	Number of surgeries (n events; [n individuals])	25 [12]
	- Headpost	1 [1]
	- Headpost removal	3 [3]
	- Microdrive base chamber	4 [4]
	- Microarrays	5 [5]
	- Craniotomy and Microdrive lowering	5 [4]
	- Procedural Intervention	7 [5]

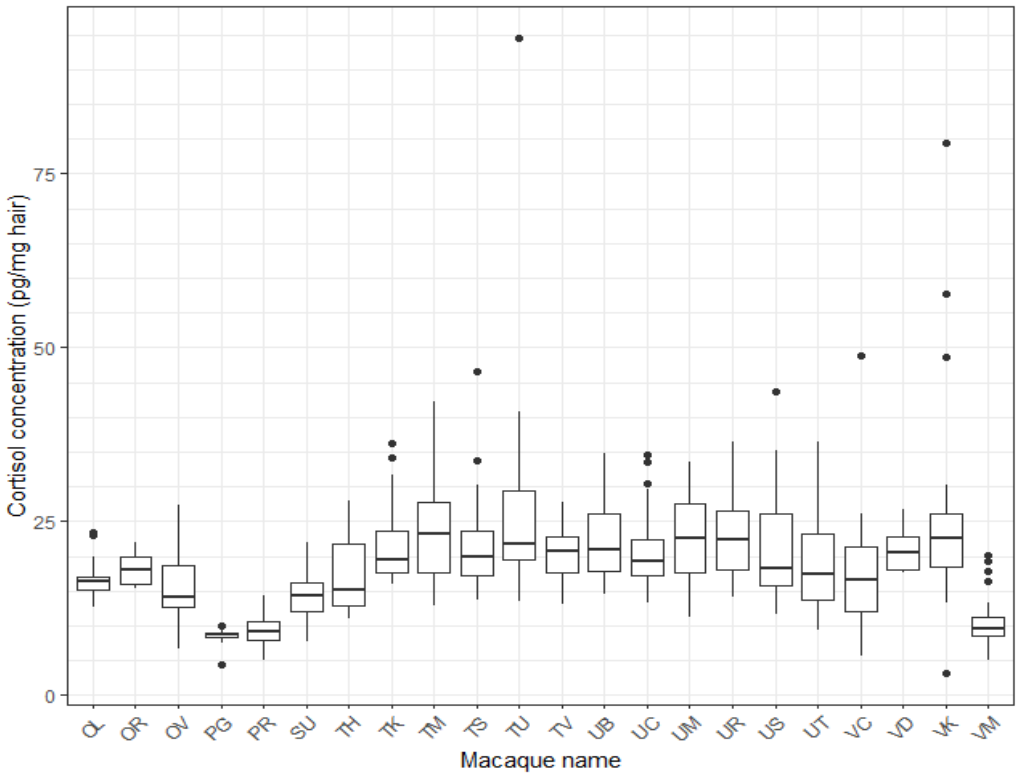

Figure 19: Boxplot (median, 1st and 3rd quartile and outliers) illustrating cortisol concentrations in hair of 22 male macaques using LC/MS analysis. Period of sample collection for each macaque varied between 3 months and 3.9 years (mean 2.3 years). All macaques were participating in neuroscience research protocols during sample collection. Neuroscience protocols for all macaques were overall similar, but not identical. Black dots represent outliers.

All samples were assayed at the same time (June 2021), therefore storage times varied from about 1 – 5 years. Results on the time of storage on HCC are presented in Table 5.3 and suggest a significant but small storage effect on these samples. According to this LMM, HCC decreases by 0.01% per day in storage (3.65% decrease per year). Inter-individual variation explained 28% of HCC variation (ICC of identity =0.28) and social group for 16% of HCC variation. Figure 20 illustrates the HCC profile of the three macaques with longest sampling collection periods (almost 4 years). There is no indication of cortisol leakage over time and cortisol peaks within individuals reach similar heights over years. A plot showing HCC over time for all macaques is provided in 'Appendix 3**'**. Here we see evidence of a subtle cortisol increase over time in 6 out of 22 macaques.

Table 5.3: LMM output looking at the influence of the time spent in storage (days) on the log of hair cortisol concentration (all data from 22 male macaques were used in the model). Significant effects can be seen in bold.

Fixed Effects	Estimate	SE	t	d.f.	p
Intercept	3.00	0.08	36.22	63.22	
Length of Storage (days)	**-0.0001**	**0.00005**	**-2.15**	**453.81**	**0.03**

Pseudo-R^2 (fixed effects) = 0.008; Pseudo-R^2 (total) = 0.44; ICC (Identity) = 0.28; ICC (Social Group) = 0.16

Figure 20 shows a peak in HCC in early 2018 and again in early 2020 for UC and UT. These animals were group-housed together (along with one other animal - VM). Life records and notes made throughout the study report the introduction of 4 new animals (2 in mid-January and 2 in mid-February 2018) into the room where UC and UT were housed. Furthermore, one male macaque who was housed in the room for 3.5 years moved out at the end of January 2018. Of the new occupants, one was a female in a previously all-males-room and overall, the group composition changed from 6 to 9 macaques. This social disruption was of indirect nature as UC and UT's social dynamics did not directly change (remained within their group of 3), but housing within the facility is such that macaques can see, hear, smell, and interact with other room inhabitants. In early 2020, UC, UT, and their other social partner (VM) all underwent the same surgical procedure. Surgeries were conducted on different days over a 1-2 week period. Following each surgery, the focal animal was separated from their social group for 1-2 days during recovery, thus the group experience a period of social disruption in addition to the surgeries.

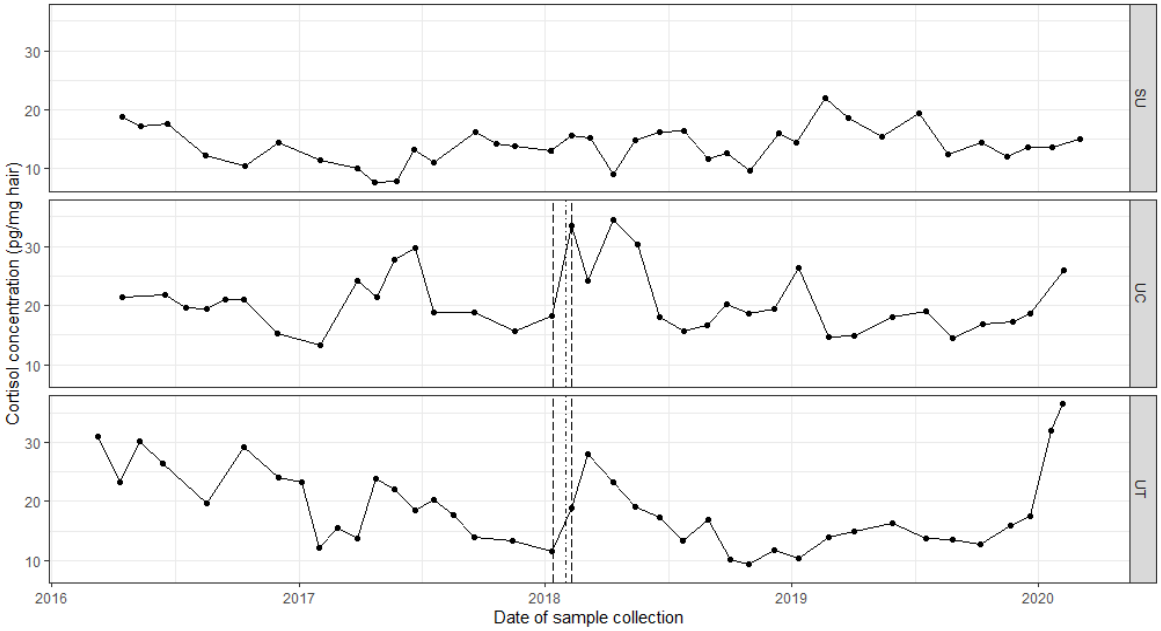

Figure 20: Cortisol concentration in hair over time of 3 male macaques (SU, UC, UT) with longest running data collection period (3.8 - 3.9 years). Lines on UC and UT plot represent a series of indirect changes to the housing of these two animals. Whilst their immediate social group didn't change, social composition of their room was altered. 2 new animals (1 male, 1 female) moved into the room (first dashed line). Then a male macaque moved out of the room (dot dash line). Finally, 2 more male macaques moved in (last dashed line).

In Figure 21 illustrates the HCC of two macaques over an 18 month period, with respect to the dosage of oral prednisolone treatment they received during that time. The 'total dosage' is the sum of daily doses given between two hair sample collections. Some of the lower doses (not dosage=0) reported consist of several days where no prednisolone was given at all. Dosage varied with symptomatology, ranging from 25mg per day to 2.5mg per day whilst on treatment. It becomes apparent that prednisolone treatment resulted in reduced HCC even at very low dosage. HCC levels after treatment ceased (dosage = 0) were comparable to other animals ('Appendix 3').

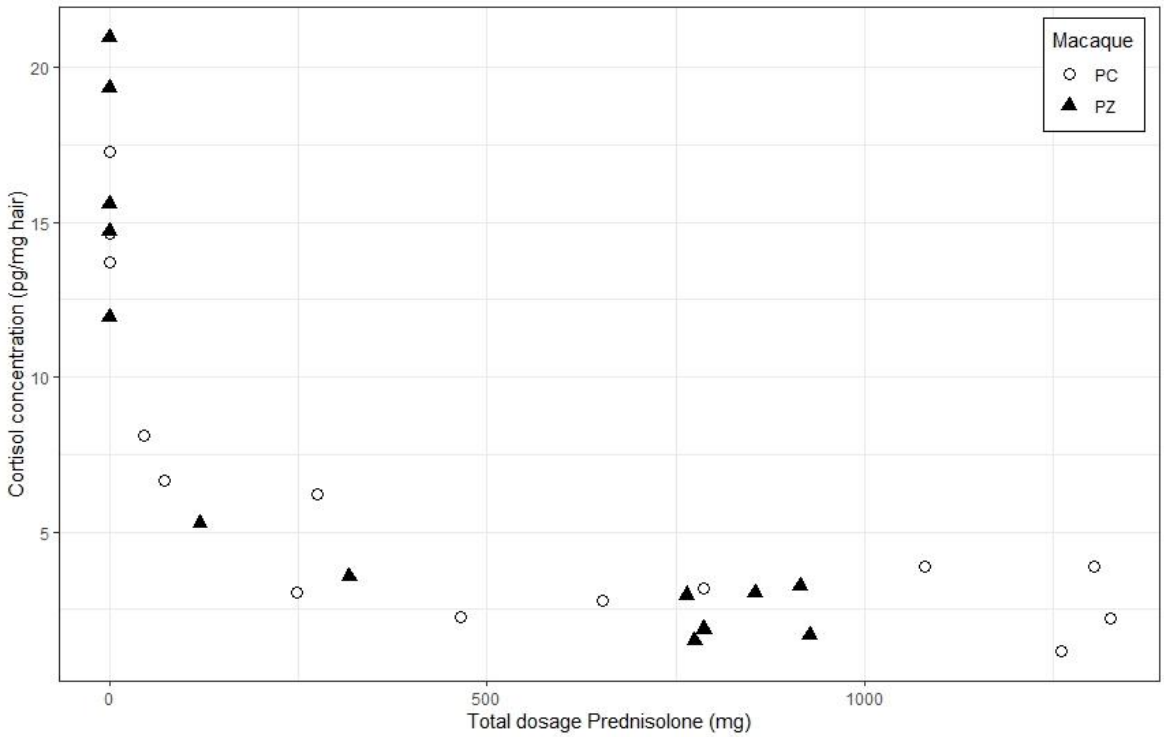

Figure 21: The effect of prednisolone treatment on the hair cortisol concentration (pg/mg) of two male macaques (PC and PZ). Total dosage (mg) represents the cumulative amount of treatment the animal received during the window between hair samples.

A LMM to assess the effects of various intrinsic and extrinsic factors was conducted on the log of hair cortisol. Results are displayed in Table 5.4. LMM shows that the only fixed effect with a significant (p=<0.001) influence on HCC was the occurrence of a surgical intervention. The effect estimate for surgery was 0.22 for the log of HCC and when transformed back, reveals a 25% increase in cortisol as a result of a surgery. Fixed effects as a whole only explained 9% (pseudo-R^2) of the HCC's but total effects, inclusive of the random effects of individual identity (animal name) and social group, explained 49% of the HCC. HCC shows a large amount of variation between individuals, with individual identity explaining 30% of the HCC values (ICC identity). Some variation is also seen between social groups, explaining 14% (ICC social group) of HCC.

Table 5.4: Outputs from the LMM looking at influence of multiple fixed effects on the log of hair cortisol concentration (pg/mg) (all data from 22 male macaques were used in the model). Significant effects can be seen in bold.

Fixed Effects	*Estimate*	*SE*	*t*	*d.f.*	*p*
Intercept	2.27	0.31	7.23	49.54	
Time In Oxford (years)	0.02	0.02	1.15	368.94	0.25
Age On Arrival (years)	0.12	0.07	1.79	32.15	0.08
Single Housing (yes/no)	0.16	0.11	1.48	14.91	0.16
GA/MRI (yes/no)	0.08	0.06	1.51	511.31	0.13
Surgery (yes/no)	**0.22**	**0.05**	**4.42**	**513.22**	**<0.001**

Pseudo-R^2 (fixed effects) = 0.09; Pseudo-R^2 (total) = 0.49; ICC (Identity) = 0.30; ICC (Social Group) = 0.14.

An outlier test was performed to check for extreme values of HCC. This highlighted 3 outliers, 3.24 pg/mg, 3.25 pg/mg (both from macaque VK) and 94.52 pg/mg (macaque TU). The LMMs were run with and without these outliers included in the dataset. Both LMMs showed the same significance effect and only slight variation between estimates, pseudo-R^2 and ICC's (intraclass correlation coefficient). Therefore, only LMM outputs including outliers are reported. LMM tables excluding these outliers can be found in 'Appendix 3'.

For single housing, GA and surgical procedures, a binary (1/0, yes/no) variable was used. There were 7 occasions (across 5 macaques) when social housing configuration changed (i.e. group housed to singly housed or vice versa) between two sampling periods. In these cases, the sample following the social change was classified based on the social grouping they experienced

on the majority of days (ranging from 51-97%) between that sample and the previous one. No macaques in this study were singly housed for their entire collection period, with the longest single housing period being 1 year and 2 months.

There were instances were multiple MRI or surgery events occurred between two hair samples. In 6 cases, macaques experienced 2 MRIs between hair sampling windows and in 8 cases, there were 2 surgical procedures. In one instance, an MRI and surgical procedure occurred between sample collections. The binary classification within the model does not account for the effect of multiple events. However, as there were few of these, model predictions from the LMM were explored and compared to the actual HCC for those events. There was no pattern observed that suggested multiple occurrences of GA within one hair sampling window caused any further effect on HCC than that exhibited by the single event.

Individual animal plots of HCC over time were created (not shown here) and explored for patterns associated with particular events. Figure 22 shows a case study of 3 macaques; all enrolled on the same study and experiencing the same protocols required for the fitting of a microdrive (see section 2.4.7 for further details). These protocols included two MRIs (under GA) and two surgeries. Firstly, each animal experienced one MRI (under GA), to create a 3D model of the skull required for an individually, customised design of a microdrive base chamber. Around 6 months later, 'surgery 1' was performed to implant this chamber. In macaques UR and VK, this was followed by a second MRI (under GA) 6-days post-operative. Macaque VC did not undergo a post-operative MRI, due to surgical complications. These complications instead required VC to have 'follow-up' assessments under GA, 2-3 months following 'surgery 1,' to undertake veterinary assessments and treatments due to infection. 'Surgery 2' involved a craniotomy and microdrive placement, which was conducted under a single GA for macaques UR and VK but under two GAs (11 days apart) for macaque VC.

Macaques all have HCC increase in the first hair sample post- 'surgery 1'. In UR, the increase is slight, and HCC remains elevated in a second hair sample. In VC, slight increase is seen following 'surgery 1' and then HCC further increases again after 'follow-up' assessments. In VK, HCC begins increasing following the first MRI but still shows an obvious cortisol peak following the 'surgery 1' too. Peak HCC following surgery in each animal represents a 1.6, 2.9, and 3.5 fold increase in UR, VC and VK, respectively, compared to the mean of all samples taken before 'surgery 1' (over 2 - 2.5 years). No further cortisol peak is observed in any macaques following 'surgery 2.' In fact, cortisol levels appear to return to within range of previous values.

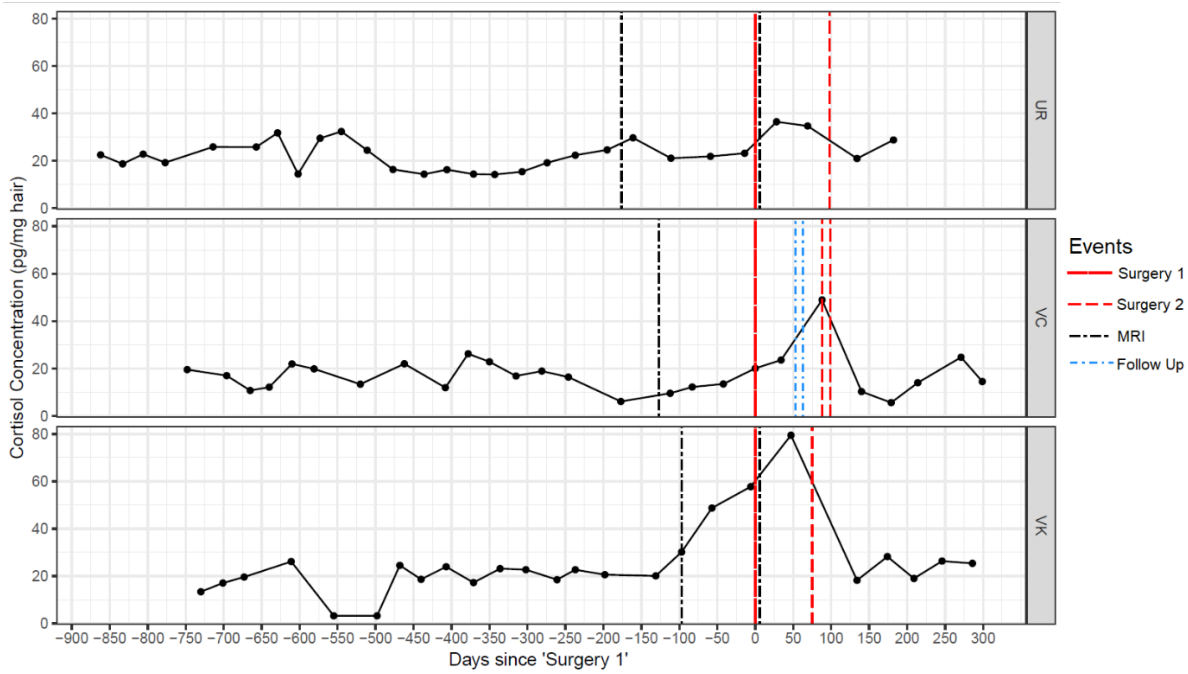

Figure 22: Hair cortisol concentration (pg/mg) of 3 male macaques (UR, VC, VK) that underwent the same surgical procedures as part of their scientific protocols. All macaques experienced a chamber (microdrive vase) surgery ('surgery 1' at day 0, MRI event >6 months prior to that surgery and a craniotomy and microdrive placement surgery ('surgery 2') <3 months following 'surgery 1' (see section 2.4.7 for further details on protocols).

5.2.4 Discussion

Measures of HCC are becoming increasing popular in the field of animal welfare as a means of the assessment of chronic stress. In this study, we have further validated its usefulness for longitudinal welfare assessment in rhesus macaques housed within a laboratory environment and engaging in neuroscience research. We have explored the cortisol range of macaques housed at the neuroscience facility in Oxford, established stability of cortisol within the hair and examined the effects of several potential stressors caused by their research careers.

5.2.4.1 Stability of hair cortisol in time

Through 1 – 5 years in storage, this analysis has shown some evidence of HCC degradation over time. Ranges of HCC levels during periods of no external stressors and concentration peaks during periods with known stressors were comparable throughout time (see Figure 20 and 'Appendix 3').

As all samples in this study were assayed at the same time, increased storage time means that a sample was collected earlier on. In turn, this means that samples stored for longer from the same individual represent that individual when they were younger and had experienced less exposure to stressors of neuroscience research. Part of the degradation effect seen could be attributed to these factors, although it was not possible to isolate these impacts in this analysis. To an extent, this experiment did control for these additional factors because individuals were added opportunistically. Some older samples were from macaques in the latter stages of long-term studies and some more recent samples were from macaques newer to the facility.

There is contrasting evidence of cortisol degradation in hair samples over time. This has been studied by comparing the cortisol concentrations in the proximal to the distal hair segment. There are reports of HCC remaining stable throughout the whole length of a hair sample (Carlitz et al., 2014; Yamanashi et al., 2013). However, other studies have reported a decline in cortisol concentration from proximal to distal hair segments (Carlitz et al., 2016; Kirschbaum et al., 2009).

However, this effect is most likely attributed to a waning effect of regular exposure to water and UV. This is unlikely to have affected influence in this study, as indoor-living minimises exposure to these elements and samples were regularly collected every 6-8 weeks.

It has been theorised that, once collected, as long as hair samples are stored away from these exposures (i.e. dry and in the dark, at room temperature) then no further cortisol degradation would occur. Some studies in mammals have agreed with this, showing no evidence of cortisol degradation during 12 months of storage (grizzly bears (Macbeth et al., 2010) and cattle (Gonzalez-de-la-Vara Mdel et al., 2011)). However, a study in Egyptian mongoose found evidence of cortisol degradation in hair samples during long term storage (2-6 years), with a negative effect observed between HCC and storage time (Azevedo et al., 2019). As cortisol degradation in hair would be expected to occur slowly over-time, it is likely that the small effects observed in the mongooses and in this study are evident because of the longer periods of storage overall. Therefore, care should be taken when interpreting cortisol from hair samples that have been in storage for several years. Whilst this does not prevent HCC from being a valuable measure, it may be that a correction to the data may need to be applied in order to minimise the effect. An even better approach would be to minimise storage times where possible, as human hair samples have been found to have reduced cortisol concentrations after a storage of just 1 year (E.H.D. Carlitz, personal communication, March 2023).

From observations in this study of individuals HCC over time, only a few animals (6 of 22) showed HCC increase during the very last year of sampling. This suggests that this pattern relates to environmental stressors instead of a common degradation process over time. It is unclear whether we see a cumulative effect of stressors in these animals, but it is interesting to note that HCC increase was only minimal. Another possible explanation for the increase in HCC would be a possible cumulative effect, where cortisol levels are ‘stacking up’ in response to exposure to multiple stressors (Pickard, 2013).

5.2.4.2 The case of Prednisolone treatment

The case of two macaques receiving prednisolone treatment has acted as a natural validation for cortisol detection in hair. Prednisolone is a synthetic glucocorticoid (GC) and will bind to GC receptors at the hypothalamus and pituitary. Due to the negative feedback effect of the HPA axis, this binding will suppress release of CRH and ACTH. Thus, in turn, suppressing the release of cortisol into the bloodstream (Dogra & Vijayashankar, 2022). There are lots of examples in the literature of cortisol suppression caused by administration of synthetic GC, particularly as corticosteroids are used for the treatment of asthma. In line with this, plasma cortisol levels have been reduced by oral prednisolone treatment (Wilson & Lipworth, 1999) and hair cortisol suppression is seen in asthmatic children taking synthetic GCs (Kamps et al., 2014; Smy et al., 2015). Thus, the low cortisol levels observed in macaques during prednisolone treatment are what was expected and provides evidence that cortisol levels from hair samples are genuine. This finding is also interesting from another perspective. It shows that prednisolone treatment blocks cortisol secretion almost entirely from a very low dosage onwards. As the HPA axis is involved in many central aspects and body functions, prednisolone should always be administered with great care and be considered when using these animals for research tasks. It is, however, reassuring, that HCC resumed levels of normal ranges after treatment was ceased. The measures from the two animals further suggest that very low HCC levels (below 10 pg/mg hair, measured with LC/MS-MS) may indicate some dysfunction in the HPA axis. Furthermore, detecting cortisol in hair may offer an additional benefit to the refinement of clinical treatment in research animals. For example, using HCC as a means of assessing the effect of steroid treatment may help with veterinary decisions about dosage and treatment plans.

5.2.4.3 HCC and surgery

The outputs from the LMM predict a 25% increase in HCC following a surgery event. Cortisol increase is an expected endocrine response to surgical stress (Cusack & Buggy, 2020). In human studies, the magnitude of cortisol response is influenced by the severity of the surgery.

Serum cortisol levels on the day of surgery increase with invasiveness, including highly, moderate and minimally invasive surgeries (Prete et al., 2018). Peri-operative cortisol levels were not elevated following minimally invasive surgeries but remained elevated for up to 7 days following moderate and highly invasive surgeries. HCC represents an average of the accumulation of cortisol over several weeks and is not likely to be impacted by brief stressors (Meyer & Novak, 2012). Therefore, HCC response to surgery in this study is most likely influenced by a prolonged cortisol elevation period, such as that seen in more invasive surgeries in humans, lasting several days or weeks. It is also most likely that this HCC increase is driven by the surgical injury and recovery period rather than the GA. This is because, according to the LMM, an MRI procedure involving a GA but no tissue trauma, does not affect cortisol levels. This reflects the results observed in rabbits, where a cortical peak in hair was observed following a 'real surgery' and not following a 'sham surgery,' which was just an application of GA (Peric et al., 2018). However, studies assessing the GC response to GA alone, using hair samples, seem to be rare. This is most likely as GA recovery is usually short (hours or days) and so studies focus on the immediate post-GA period, for which a different cortisol matrix (e.g. faeces) is more appropriate. Additionally, the frequent and regular exposure to GA is uncommon outside of the field of animal research. Therefore, there may be a narrow field where this method is suitable. One study looking at laboratory mice used both faecal and hair cortisone measures, amongst others, to determine whether multiple exposure to GA resulted in chronic stress (Hohlbaum et al., 2018). They found elevated faecal cortisone following the first GA event but no change in hair cortisone levels at any point, suggesting there was no cumulative effect to GA procedures. Based on these findings and the human literature, future analysis separating surgical interventions into severity categories could to beneficial to the understanding of the causes of HCC elevation. This could highlight potential areas of refinement within surgical procedures.

A case study presented in Figure 22 illustrates a rise in HCC following a surgical intervention. This case study is taken from 3 macaques, all of which were partaking in the same

scientific protocols. Although of varying magnitudes, all macaques show increased HCC following a cranial chamber surgery ('Surgery 1'). This is consistent with the findings from the LMM. However, in the case of animal VC, peak cortisol levels appear in the second sample following surgical intervention, rather than the first.

Interestingly, no cortisol peak was observed following the craniotomy and microdrive surgeries. However, unlike the chamber surgeries, post-surgical treatment for craniotomies included steroids, specifically dexamethasone. As for treatment with prednisolone, dexamethasone suppresses cortisol release. Dexamethasone is a potent corticosteroid, 5-6 times more effectively binding to GC-receptors than prednisolone. A 'dexamethasone suppression test' is commonly used to assess the proper functioning of the HPA axis, by detecting the expected suppressed cortisol levels (Dogra & Vijayashankar, 2022). It is frequently used as a physiological validation for cortisol detection in biological samples (Palme, 2019). Steroids were given for up to 2 weeks post-surgery. This treatment most likely masked any stress related effect of the surgery.

Upon further investigation, Dexamethasone treatment was given after 10 of the 25 surgeries from the LMM; this included 5 array surgeries and 5 craniotomies. Both of these surgeries are highly invasive, involving exposure of the brain, and so the anti-inflammatory effects of steroids were required to reduce swelling. This finding inflates the influence of surgery on HCC seen in the LMM. As 40% of the surgeries used in that prediction were likely to be influenced by dexamethasone suppression, still observing a 25% increase in cortisol suggests that HCCs are much higher following the remaining surgeries.

In the case of macaque VK, we observed HCC increasing prior to surgical intervention. The increase begins following the MRI event. As seen in the rest of the data, MRI's associated with GA do not appear to drive cortisol increase. However, in this particular case, macaque VK experienced social distress before the MRI. More specific, VK and his social partner had a social dispute that culminated in a fight, during which VK was injured a few days prior to the MRI.

Attempts to re-pair the two after the MRI were unsuccessful and the decision was made to separate the macaques permanently. Therefore, the increasing HCC before the first surgery most likely results from a cumulative effect of the social stress, rather than from the MRI under GA. Interestingly, surgery 1 poses an additional stressor resulting in an even higher HCC peak. The latter aspect is interesting as it shows that there is no ceiling effect in this range and could be evidence of a potential 'stacking up' effect of many stressors occurring without sufficient recovery time in between (Pickard, 2013). Hair cortisol in rabbits showed a similar cumulating effect, with elevated cortisol following a relocation being compounded by a surgery soon after (Peric et al., 2018). Using HCC to identify such cumulative effects could help to refine protocols and optimise animal welfare. For example, knowing which events may have a cumulative effect allows for better management of scientific methods and avoidance of conducting surgeries too close to additional stressors.

In summary, the HCC measures show that surgeries in the research facility are a severe stressor to the animals, but the HCC measures from three examples (Figure 22) suggest that the animals do fully recover over time from an endocrine perspective. However, the example of macaque VK underlines that the influence of social stressors may compound this effect if they occur in close proximity.

5.2.4.4 Single housing and social stress

Interestingly, single housing does not seem to influence HCC in this dataset. A lack of socialisation opportunities is considered to be a major psychological stressor for NHPs (Rennie & Buchanan-Smith, 2006b) and the literature about the stress of single-housing in macaques is vast (DiVincenti & Wyatt, 2011). The lack of impact on cortisol in this study could be due to small sample size, with only 5 of the 22 macaques experiencing single housing and of those, none were housed alone for the whole collection period. Furthermore, singly housed animals in this study are not devoid of all social interactions. The housing set-up at Oxford facilitates indirect social

interactions with other macaques housed in the same rooms. In line with this, a study assessing plasma cortisol in rhesus macaques found no difference in cortisol between singly and pair-housed individuals (Schapiro et al., 1993). They did, however, find cortisol difference between indoor and outdoor housing, suggesting that the environment outside the cage may have an influence on animal stress.

In addition, group or pair housing under confined conditions may provide serious stressors, too. Competition for dominance or mates, arguments with conspecifics and individual rank are all stressors common within primate societies (Sadoughi et al., 2021). Rhesus macaques have a despotic social style and maintain their social systems through frequent and, often, severe aggressive interactions (Thierry, 2007). In captive housing, particularly within a laboratory, restrictive living space can result in more forced social interactions between group mates or an inability to engage in avoidance behaviours. This may increase levels of both, psychological and physiological stress. Instability within a social group increases agonistic behaviours, such as threats and chasing. Studies comparing stable and unstable social groupings in primates have found conflicting evidence in the HPA axis. In macaques hair, hypocortisolemia and hypercortisolemia is observed in unstable social groups (plasma cortisol (Capitanio et al., 1998) and hair cortisol (Vandeleest et al., 2019), respectively). Unstable groupings within a laboratory environment are likely to result in fighting between conspecifics and alterations in housing arrangements, e.g. separating or regrouping animals. These social housing changes can have an impact on behaviour and physiology indicative of diminished welfare, particularly when reducing group sizes, which may be short or long lived (Hannibal et al., 2017). Under these conditions, it is not surprising that we did not find differences in HCC between single and group housing.

Looking at individual HCC profiles, this study observed some cortisol peaks that align with periods of social instability. Animal VK experienced an MRI procedure under GA, a fight with his social partner and subsequent separation from that partner within a few days. Following this

series of events, HCC levels started to increase (Figure 22 at day -100). Furthermore, social disruption caused by the addition of several new macaques into a holding room may have contributed to elevated HCC levels in two group-housed macaques in 2018 (UC and UT from Figure 20). A return to the HCC levels as they were before took several months. A similar pattern is seen in other animals who were also housed in this room during the disruptive period (VM and TV, see 'Appendix 3.') This is an interesting and important observation because these social stressors may influence the results from the research experiments for which the animals are kept. Thus, it is advisable that group composition within a room should be held as constant as possible.

These results highlight the influence of social stressors on animal welfare. It would be beneficial to understand the impact on welfare of different social experiences. This could be achieved using the HCC data from this study and animal life records to identify social events and categorising them (e.g. gaining or losing a social partner). Quantifying the stress caused by such experiences could help to make decisions about approaches to social housing which would maximise animal welfare.

5.2.4.5 Inter-individual variation

In this study, a high degree of variation was observed between individual's HCC measures. HCC is thought to offer a stable, within individual measure of cortisol. That is compared to other mediums, such as blood and saliva, which express a high degree of intra-individual variation when used for long-term cortisol estimation (Stalder et al., 2012). Inter-individual variation is often under-reported in cortisol studies but there is evidence that the degree of response varies extensively between individuals (Cockrem, 2013). Exposure to the same stressor can evoke very different cortisol responses in different individuals.

Whilst the source of this variation between individuals is difficult to identify, potential contributors include differences in genetics, previous stress experiences, and personalities (Cockrem, 2013). This extensive inter-individual effect suggests that the cortisol response to some

stressors could be much more severe in some animals compared to others. In terms of animal welfare, this implies that some individuals are better able to cope with certain neuroscientific procedures. Therefore, to understand the true welfare impact of particular procedures, comparisons within individuals are extremely important.

To a lesser extent, variation in cortisol response between social groups was observed in this study. For ease of facility management, social groupings are often created such that macaques undertaking similar or the same neuroscience protocols are housed together. Furthermore, some social stressors are likely to affect the social group as a whole. Therefore, socially housed animals are likely to have more similar experiences to their conspecifics than they are to animals housed in different groups and so cortisol responses would predictably be more similar to that of their conspecifics too. Additionally, the stability of social groups can influence hair cortisol in macaques (Vandeleest et al., 2019) and this could also play a role in driving the variation observed between groups.

5.2.4.6 Conclusions

Measures of HCC have demonstrated their usefulness as a non-invasive measure of chronic stress that can be easily adopted alongside neuroscience research. HCC are able to detect a response to severe or sustained stress in the form of surgical procedures or social stress. However, HCC does not appear to be influenced by more transient stressors, such as an MRI conducted under GA. Measuring the activity of the HPA axis is not always suitable for measures of surgical stress, as post-operative treatments result in HPA axis suppression, masking the stress response.

The variation in HCC between individuals is large and so it is invaluable to track animals longitudinally, using within individual comparisons for ascertaining the welfare effect of a life in neuroscience research.

This study has highlighted the influence of social stress on HCC. Macaques experiencing either direct or indirect changes to their social dynamics appear to elicit a stress response of sufficient magnitude to be detected in hair samples. This emphasises a need for careful consideration of NHP social groupings and the maintenance of a stable social environment. Furthermore, when social stress occurred in close temporal proximity to other stressors, this study showed some evidence of a potential cumulative effect. Although this sample size is small and requires further investigation, it may still highlight potential for refinement. For example, by being able to be flexible with the timings of experimental protocols, researchers could avoid conducting surgeries during periods of social stress or instability to minimise cortisol increase.

Understanding cumulative experience in research animals opens up avenues for welfare refinements. Welfare compromise in research animals is inevitable, as many methods considered essential for successful scientific outcomes are going to cause harm. However, by monitoring cumulative experience, we can hopefully highlight any unnecessary harms to minimise suffering.

5.3 Overall discussion

Non-invasive methods of cortisol assessment may offer a valuable contribution to the understanding of the welfare of research animals. The ability to quantify the stress response to various scientific practices may highlight areas for future refinements and result in improvements in animal wellbeing. Optimal methods of welfare assessment in research animals are those that cause minimal disruption to the animal's wellbeing, whilst also avoiding disruption of the scientific goals. Furthermore, methodologies that require little time and resource investment are the most appropriate for encouraging busy researchers to run welfare studies in parallel to their own research. After trialling 4 different sampling methodologies, it was apparent that collection of faecal and hair samples for cortisol assay were the most ideally suited for this environment. A further benefit of these sampling methods is that they lend themselves well to longitudinal data collection. This associates well with NHP and neuroscience research, where studies usually last several years.

Both HCC and FCM were considered in response to a general anaesthetics. In faeces, there was an increase in FCM, indicating an acute stress response, to both an MRI and a surgical procedure. HCC did not appear to be effected by an MRI event but was elevated following a surgery. A similar finding was observed in mice, where cortisone metabolites in faeces were elevated in response to single anaesthesia but no such effect was seen in the hair (Hohlbaum et al., 2018). Generally, HCC is considered a useful tool for measuring long-term or chronic stress, as transient stressors seem to go undetected (Heimburge et al., 2019; Meyer & Novak, 2012). This is because HCC reflect HPA axis activity over the several weeks of hair growth, so short, infrequent changes to cortisol do not drive the integrated average. These results suggest that administration of a GA alone is mild, transient stress but a surgical procedure will have a much more profound impact on chronic stress. This falls in line with what was expected and what is currently reflected

in the law (Animals (Scientific Procedures) Act, 1986), where an MRI is classified as a 'mild' severity procedure and a surgery as a 'moderate' or 'severe' procedure.

The surgical procedures assessed using faeces were not the same as the procedures assessed using hair cortisol, so the two measures are not completely comparable. The opportunistic data collection methods employed here were the main reason for this. For the FCM dataset, one of the reasons for assessing only headpost surgeries were for consistency and these being the most commonly conducted procedure within the facility. However, for the hair sampling methodology, a headpost must have already been fitted before animals could be engaged on studies, as they were required for the appropriate restraint. As fewer occurrences of all other surgical procedures took place, splitting hair cortisol measures by specific surgical procedure would have been limiting. Grouping together similar surgical events, e.g. by level of perceived severity, could offer some interesting insights and is a potential avenue for further investigation.

Surgical procedures often require post-operative steroid treatment, to avoid inflammation. The findings in this study match the expectation that the administration of such exogenous GCs severely suppresses the HPA axis activity and reduces the endogenous GC levels (Dogra & Vijayashankar, 2022). Therefore, assessment of cortisol concentration, using any sample medium, following a surgical procedure may be an unreliable and inappropriate measure of stress.

Cortisol responses in this study were elevated in response to stressors. This reflects the results found in many other studies. However, hypocortisolemia has also been observed (Capitanio et al., 1998; Fries et al., 2005; Steinmetz et al., 2006). In which, baseline cortisol values are lower and stress responses are suppressed due to the chronic stress exposure depressing the activity of the HPA axis. These two opposing outcomes can make identification of a stress response quite difficult by using cortisol measures alone.

Measuring cortisol in hair and faeces offers the opportunity to monitor both the magnitude and duration of the stress response. Total exposure to GCs is what will effect biological processes and overall welfare, whether that is through a brief but large response or a sustained, weaker response (Sheriff et al., 2011).

Although data for this study were collected opportunistically, collecting from a laboratory environment means that some of the factors that could potentially influence cortisol are relatively controlled. Some of these factors are age, sex, social rank, and individual differences (Palme, 2019). For example, all macaques in this study were adult males, so no effect of sex or age is likely to be observed. Factors such as social rank could be further investigated, as extensive animal records are kept for laboratory macaques and this information is available ad-hoc.

Many parts of the HPA axis are susceptible to individual variation. These differences could be caused by genetics or previous life experience (Palme, 2019). However they come about, their influence must be carefully considered when using GCs as a marker for stress. Both hair and faecal cortisol assessment lend themselves well to longitudinal application. This offers the advantage of animals being able to serve as their own control and set their own baseline (Palme, 2012). In this study, where no controls were available, both of these methods were ideally suited. Furthermore, these methods offer the opportunity for animals in other studies to be their own control. Thus reducing the number of animals needed and helping to satisfy the 3R requirement in laboratory animals (Russell & Burch, 1959).

The methodologies employed in this study worked well for this environment, with generally good communication and collaboration between the welfare researchers, neuroscience researchers and animal care staff. There are some elements of the methods that should be carefully considered if these were to be applied elsewhere. Sample contamination is a concern with this methodology but ensuring that personnel are appropriately trained should enable optimal collection protocols to avoid this where possible. Hair samples were collected under

restraint, which is inappropriate in some environments or at least requires the suitable licenses and training for laboratory animals. Some storage space is required for longitudinal collection, particularly for faecal samples, that should be stored at -20°C. Although still simple collection methods overall, both hair and faecal sampling do require some resource and time investment.

Whilst there are many benefits to using GCs in animal welfare studies, it does need to be approached with caution. A major challenge of GC assessment is that stress is not the only stimuli that results in increased GC activity. Cortisol increase has been observed in response to increased physical activity, sexual behaviour, and anticipation of reward or punishment (Dawkins, 2006). HPA activity increase is likely linked to physiological arousal, which could be either a positive experience (e.g. sexual behaviour) or negative (e.g. stress) (Otovic & Hutchinson, 2015). In this study, we can assume from context that the administration of a GA is probably a negative experience. Thus, the elevated cortisol levels observed are most likely caused by stress. However, the same assumptions could not be made about other neuroscientific practices we have discussed. For example, animal training sessions could be considered negative, due to the use of restraint, or positive, offering a form of enrichment. For ambiguous experiences such as these, a secondary welfare parameter should be employed as a bare minimum, ideally one to indicate emotional valence.

Due to the collected samples already available and the simplicity of adopting these methodologies again, there is a huge amount of potential for further study. This could be to assess the stress response to numerous other neuroscience experiences or investigate the influence of certain factors, such as social rank, on cortisol concentration.

Whilst some studies have explored both hair and faecal cortisol, (Dulude-de Broin et al., 2019), there is a lack of direct comparison. Although there is some evidence that both measures reflect the same adrenal activity (Accorsi et al., 2008), applying the methods from this study in parallel could help to establish if hair and faecal cortisol tell the same story. Furthermore, using

both measures longitudinally, a full life history of an individual animal's acute and chronic stress could be obtained. This would be particularly valuable for assessing the impact of repetitive stressors. Due to a lack of full control over sampling opportunities, there was no scope to achieve these comparisons using our dataset.

It is apparent that using faecal and hair cortisol as a welfare assessment offers an opportunity for the implementation of a simple methodology, which works alongside scientific research protocols. Arguably, there are welfare parameters that provide fewer confounds and are better suited to monitor animal welfare (Novak et al., 2013). However, there are many reasons why these may not be suitable for laboratory housed macaques. Interpretation of a single cortisol measure may be ambiguous, particularly without context. Therefore, the best approach for welfare assessment would be to use cortisol measures alongside other parameters. Some of these may already be conducted daily, as part of macaques monitoring and cage-side assessments.

The application of more and new techniques for welfare assessment is a key component for understanding NHP welfare and for being able to apply the 3Rs. What is needed is a clearer picture of the life-time experience of the macaques engaged in neuroscience studies. Until such a picture is formed, implementing refinements will pose a continual challenge. Getting those involved with the neuroscience research, including investigators and care staff, could be greatly beneficial to both adding to this evidence and for promoting the willingness to implement change (Lemon, 2018; Prescott et al., 2017). The assessment of cortisol, in both hair and faeces, would add to this life-time experience picture and encourages direct involvement from macaque care and research staff.

Chapter 6 - Development of a novel ECG device for measuring heart rate variability

6.1 Introduction

Stress is a key contributor to the overall welfare of animals. In addition to activating a neuroendocrine response (discussed in 'Chapter 5 -Cortisol and non-human primate welfare') stress also activates a response in the autonomic nervous system (ANS) (Moberg, 2000). The two counteractive components of the ANS, the sympathetic (SNS) and parasympathetic (PNS) nervous systems, regulate many biological processes including the cardiovascular system. In response to stress, the balance between the SNS and PNS shifts leading to alterations in these processes, e.g. increased heart rate (HR), in order to restore homeostasis (Kim et al., 2018; Moberg, 2000).

Monitoring HR does reflect the activity of the ANS and can be used as a measure of the stress response. However, by measuring HR alone the interplay between the SNS and PNS is not always comprehensible. A rise in HR can be mediated by either an increase in sympathetic activity, a decrease in parasympathetic activity or simultaneous change in both systems (von Borell et al., 2007). Heart rate variability (HRV) refers to the variation in time intervals between consecutive heartbeats (Karim et al., 2011) and better reflects the sympathetic-parasympathetic balance that is regulating cardiac activity. Depending on the HRV parameters used, it is possible to simultaneously monitor SNS and PNS activity separately or to assess the balance between the two (Kovacs et al., 2014). There are many metrics that can be used to evaluate HRV and the quantity of data required is dependent on which metrics are to be applied. Longer recording periods of 24 hours offer the benefit of encapsulating the response in HRV to circadian rhythm and other physiological processes, such as sleep and metabolism. However, short recordings (1-5 minutes) can also be a valuable reflection of HRV. Most commonly, time domain (time period between successive heartbeats) and frequency domain (distribution of power into frequency

bands) measures are used to assess HRV. Selection of the appropriate domain and metric is dependent on the question being asked. For example, a time domain measure such as RMSSD (root mean square of successive differences between normal heartbeats) provides an estimate of the PNS mediated changes in HRV and in the frequency domain, whereas a Low frequency/High frequency ratio can assist in identifying whether the PNS or SNS is dominating HRV regulation (Shaffer & Ginsberg, 2017). Overall, HRV reflects the heart's adaptability in changing environments: a higher HRV suggests better flexibility and a better ability to cope with unpredictable circumstances; low HRV (i.e. a monotonously regular heart beat) suggests poor homeostatic control and a reduced ability to cope with stressors (Kim et al., 2018; von Borell et al., 2007).

Exposure to psychological and physical stress has been associated with a reduction in HRV (Karim et al., 2011). HRV monitoring in laboratory housed macaques has reported decreased variability in response to stress. Social status in monkeys appears to effect HRV measures with subordinate animals having lower HRV than their dominant conspecifics (Shively et al., 2020; Silverstein-Metzler et al., 2022). A reduction in HRV was also observed upon exposure to a novel environment (Shively et al., 2007). In these studies, chronic stressors reduced basal HRV and stress resilience. An exaggerated HRV response was observed following these acute stressors in animals consuming a poor diet or alcohol. In addition to measures of stress, an increase in HRV following a positive experience, i.e., grooming from familiar humans, has been observed in macaques (Grandi & Ishida, 2015; Grandi et al., 2015). Furthermore, HRV has been used to identify emotional states. In female cynomolgus macaques an increase in cardiac vagal activity, measured using HRV parameters, correlates with depressive behaviours (Jarczok et al., 2018). Identifying the benefits of such experiences provides great opportunities for implementing improvements to animal welfare.

HRV is typically measured by calculating the distance between consecutive R-waves from an electrocardiogram (ECG) trace. These distances can be referred to as an inter-beat intervals (IBI), RR intervals or NN intervals (represents normal R peaks, abnormal R waves are removed) (Karim et al., 2011).

An ECG measures the electrical activity of the heart and is the gold standard method for measuring HRV, providing the most reliably accurate data. A standard ECG device usually requires attachment of multiple electrodes to the skin's surface, from which wires protrude, restricting the mobility of the user. Wearable, portable devices that measure HRV are preferred for animal welfare studies. Due to this, it is common for studies to use pulse rate variability (PRV) as a substitute for HRV. Monitoring blood circulation to obtain PRV can be achieved using light that emits into the tissue. This technology is called photoplethysmography (PPG) and is popular because of its simplicity, being easily and readily added to many wearable devices such as smart watches (Castaneda et al., 2018). PPG has been used to assess pulse rate in rhesus macaques (Unakafov et al., 2018) and as an alternative to an ECG in great apes (Cloutier Barbour et al., 2020). Whilst HRV and PRV are closely related and derive from the same source, PRV is influenced by information from other physiological sources. Therefore, it is better to recognise PRV separately, instead of as a surrogate HRV measure (Yuda et al., 2020). Furthermore, measurements taken using a PPG are vulnerable to movement making them unsuitable for use in freely moving animals (Castaneda et al., 2018).

When an ECG has been used in previous NHP studies the subjects either need to be anaesthetised (Ueda et al., 2019), restrained (Grandi & Ishida, 2015; Grandi et al., 2015), implanted with a telemetry device (Silverstein-Metzler et al., 2022), or fitted with a restrictive jacket (Uchiyama et al., 2007), some of which required sedation for fitting (Shively et al., 2020; Shively et al., 2007; Ueda et al., 2019). All of these previous methods are potentially disruptive or invasive with the potential to confound results and impact animal welfare. A joint working group

of professionals in the fields of animal welfare and animal research created a report to outline refinements to telemetry procedures conducted on animals, including NHPs. They discuss that the ideal telemetry system would be minimally invasive and could not be manipulated by conspecifics, allowing animals to be housed in compatible social groups (Hawkins et al., 2004; Morton et al., 2003).

Chest worn ECG devices, primarily designed for sports medicine research, are popular in farm animal studies. These are belts or straps that fit around the thorax, in which sit electrodes that record the HRV and the resulting data is wirelessly transmitted for storage and analysis (von Borell et al., 2007). For example, these devices have reported on the HRV of sheep (Stubsjoen et al., 2015; Turini et al., 2022), cows (Kovacs et al., 2015) and outside of the farm animal field, dogs. (Jonckheer-Sheehy et al., 2012).

In NHP species chest devices would not be suitable. Due to their dexterity and opposable thumbs, NHPs could easily manipulate the ECG. This could potentially cause injury, damage to the device, or at least disruption of the data. Furthermore, the ECG of socially housed primates would also be at risk from their conspecifics too. For this same reason, measurements of PRV using PPG devices are also limited, as they must be placed at certain body locations (Castaneda et al., 2018). Therefore, the effective design of a wearable ECG device for macaques requires it to be fitted in a position where it will receive minimal interference.

Obtaining measures of HRV on freely moving animals presents further limitations. Both muscle electrical activity and movement produce signals in the same frequency as ECG and may be of equal or similar magnitude (Keith Simpson, personal communication, February 2023). This means that an ECG signal may be obscured by either of these signals on an awake, moving animal. Muscle activation results in the interference of EMG (electromyography) noise into the ECG signal, often obscuring the PQRST complex and appearing as a flicker on the baseline.

Movement can result in a wandering baseline effect, causing the shift of the ECG from its normal base (Kher, 2019).

The aim of this study was to develop a novel, wearable ECG device suitable for laboratory housed rhesus macaques. The optimal device would provide a reliable measure of HRV from freely moving, socially housed NHPs whilst causing minimal disruption to their research or welfare. This would fit the ideal criteria proposed by the joint working group on telemetry refinements (Hawkins et al., 2004; Morton et al., 2003). Furthermore, a device that was simple and easy to fit would appeal to busy researchers, encouraging ongoing data collection alongside their primary research. For these reasons, incorporating an ECG device into a neck collar was the objective of this study.

6.2 ECG development and technical barriers

In the late 19th century, Augustus Waller discovered that the heart's electrical activity could be detected by placing electrodes onto the skin (Permar & Williamson, 2021). Willem Einthoven then went on to advance these findings, creating the first practical ECG device. The standard Einthoven's triangle represents 3 planes that surround the heart; Lead I, II and III. Placement of electrodes in this formation on the body provides optimal signal detection. The outputs received from any of these leads represents the magnitude and direction, relative to the electrodes, of the heart's electrical signals. Any electrical signal observed from an ECG is the difference between two electrodes. For example, Einthoven's Lead-II is the signal between the right arm and left leg in humans (right forearm and left leg in NHPs) (Richig & Sleeper, 2019). For measures of cardiac rhythm and HRV, a rhythm strip from a prolonged reading of one ECG lead is required (Meek & Morris, 2002). This is usually created from Lead II, as this most accurately follows the natural conduction pathway of electricity through the heart and thus provides the best view of the P and R waves (Keith Simpson, personal communication, February 2023).

The small potentials gained from an ECG represent a diminished version of the collective action potentials occurring in the heart itself. These action potentials spread out across the body as an electrical vector. The ability of the ECG to detect these signals is dependent on the potential's magnitude and how the placement of the device interacts with the direction of the electrical charge. Therefore, a potential will exist between any two points placed in the path of the electrical signals, anywhere on the body. However, the magnitude of the signals will diminish the further away from the heart that the sensors are placed (Keith Simpson, personal communication, February 2023).

Whilst a standard ECG is usually fitted around the chest, this was not a suitable location for data collection from freely moving NHPs. Sensors and wires around this area would be too easily manipulated by the macaques, who would most likely be pulling and scratching at the device. Knowing that it is possible to achieve an ECG signal from anywhere on the body, it was postulated that a signal may be detectable by means of electrodes placed around the neck. Having used neck collars on macaques previously, this seemed to offer a better alternative as they seemed to habituate to them quickly and tolerated them well (see section 2.1). Furthermore, attachment around the neck is quick and simple, a method which can be performed without the need for anaesthesia and with minimal disruption to the animal. To our knowledge, our device would be novel as an ECG that can be fitted around the neck, allowing for long-term recordings that have not previously been possible to explore.

Although this placement would be ideally suited to laboratory housed macaques, there are three main consequences of using the neck to detect an ECG signal that must be overcome:

1. The detectable signal is likely to be small compared to a standard ECG due to the extra distance from the heart.
2. The presence of hair on the animal may impede signal detection.
3. The signal could be subject to movement and muscle electrical activity artefacts.

Before the design of the ECG device could be adapted for use in freely moving animals, we first needed to obtain proof of concept. We wanted to identify if an ECG signal could be detected around the neck and whether that signal was clear enough to identify RR intervals. Furthermore, we assessed whether the positioning of the sensors on the neck effected the signal and how the signal compared to a standard ECG device. In order to do this, we conducted several trials on macaques under general anaesthesia (GA) or under restraint. As the device is completely non-invasive, we were also able to test it on humans (ourselves), dogs (both awake and under GA) and a sheep (under GA).

6.3 Methods

As with other methodologies applied for this DPhil project, data from macaques were collected opportunistically around their neuroscience research. Use of the ECG device aimed not to interfere with either the scientific protocols or the animals' welfare. For this reason, ECG trials under restraint were conducted at a time when the macaque was head-fixated within the primate chair as part of their regular neuroscience testing regime. Trials under GA were conducted in line with scientific protocols requiring GA (e.g. surgery or MRI). Limited time was available to obtain the ECG trace under GA, as trials were performed during prep or whilst wrapping up prior to extubating. This avoided disruption of the scientific protocol and prolonging anaesthesia for the sole purpose of ECG data collection. When collecting data from other animals under GA (dogs and sheep), prolonged anaesthesia was also avoided but the ECG trial could be maintained throughout the whole GA if needed.

6.3.1 The design of the neck-based ECG

The neck-based ECG device consisted of the following elements. Three sensors (green, red, and black) were attached to a Velcro wrap which was placed around the neck of the subject. When running trials, two different types of sensors were tried. Limb plate electrodes, which were

flat, stainless-steel plates (3 x 2cm) that threaded through the Velcro wrap and attached to the ECG cables via a 4mm banana plug. Alternatively, the customised pronged sensors (see section 6.3.2) were tried, which were made of silicon and approximately the same size as the plate electrodes, connected to cables via a snap style plug. A conducting medium in the form of a gel (Signagel electrode gel or Aquasonic ultrasound gel) was applied when necessary to capture an ECG trace. A single-lead Sapphire plus system was used, with a transmitter connected to the ECG cables sending the signal to a receiver, which was plugged into a laptop via USB. VitalMonitor software was used to view the ECG trace in real-time and VitalStore used to observe and analyse signals post-hoc. All elements were manufactured by Vetronic Services Ltd (now Burtons Medical) and available for commercial sale, except for the customised silicon prongs.

6.3.2 Detecting an ECG trace from the neck

Firstly, this study aimed to determine whether an ECG signal, with discernible R waves, could be detected around the neck of a rhesus macaque. To achieve this, adjustments to the ECG device were made during the development stage to overcome the challenges 1 and 2 listed in section 6.2). To address challenge 1, these adjustments included the implementation of a 3-point sensor design and the incorporation of special-purpose amplifiers.

Electrical signals detected by a standard ECG are often quite small, requiring amplification to be observed and assessed. For example, the standard R-wave detected using Lead-II on a dog has a mean height of less than 2mV (Mukherjee et al., 2020). An amplification factor is applied to signals and, for a 2mV reading, typically an amplification of around x2-3000 is required to make it suitable for use in standard ECG device. The magnitude of a signal detected around the neck is much smaller and may be dependent on positioning.

As expected with a chest ECG, the greatest signal from a neck-based ECG would be when placed across the heart, i.e. left to right across the neck. However, the orientation of the heart and a ‘funnelling’ of the signals up the neck may lead to alternate configurations producing the

optimal signal. This effect was observed during testing phases of the neck ECG in humans. The resolution to this was to use 3 electrodes, in a triangular formation. This would allow for the detection of the larger signal from any one combination of two electrodes (1-2, 2-3, 1-3). Due to wanting to keep the ECG as unobtrusive and comfortable as possible, the sensors would not be attached to the neck, instead just sitting between the collar and the skin. Therefore, movement around the neck during wear is likely. As positioning could not be maintained, the use of 3-point sensors would allow for two sensors to always be able to detect the optimal signal.

This 3-point design aims to detect the best signal possible. However, that signal is likely to still be very small and would require extra amplification to be readable. At higher levels of amplitude, noise is a definite consideration. As part of the design of our neck-based ECG, special-purpose, low noise differential amplifiers were used to minimise input noise. Tight filtering (0.5 to 50Hz) was used in all gain stages followed by a final processor-adjustable filter to obtain the best quality signal. The results from this were a relatively clean signal (Keith Simpson, personal communication, February 2023).

To address challenge 2, specialised pronged sensors were devised made from conductive rubber.

The optimal ECG signal is obtained when there is minimal resistance between the skin and the ECG sensors. The normal, outer, keratinised layer of skin creates resistance with the ECG electrodes but is substantially reduced by application of a conducting medium (ECG gel) (Keith Simpson, personal communication, February 2023). A further barrier to obtaining a good signal is the presence of hair or fur, which forms a high resistance path for the signals. For clinical ECGs, either in humans or animals, hair is often clipped or shaved to remove this barrier. As one of the objectives for developing this device was to allow long-term monitoring, applying a conductive medium would not be appropriate. Most media would dry out or evaporate over time, which would create a further barrier and lead to increased resistance. Furthermore, clipping or shaving

the animal would not be suitable for this study. A large area would need to be shaved, as the collar can move about the neck, and there would be concerns about skin irritation and discomfort.

Pronged sensors resembled a small hairbrush, and each prong was approximately 10mm in length, spaced 5mm apart on a 4 x 5 grid. Placed under the collar, the prongs would allow gentle and persistent penetration through the hair to contact the skin underneath. The large number of prongs increased the area of contact and reduced contact resistance. Following the inevitable movement of the collar, the sensors were designed to be mainly unaffected. Furthermore, the sensors are designed such that they could be easily fitted inside a collar, fixing the position of the sensors to obtain the optimal signal.

Once developed, the ECG was trialled on 8 macaques whilst under GA and 2 macaques who were restrained (in line with neuroscience protocols). When conducting trials to test the ECG device, specialised pronged sensors were not always applied and instead limb plate electrodes were used. This was due to time constraints and prioritising obtaining an ECG signal, as prongs were experimental and less reliable than limb plate electrodes on immobile animals. However, development and testing of the pronged sensors was important for the future of the ECG device when sensors may move about the neck.

6.3.3 Positioning of the sensors

It was important to distinguish how the ECG signal would vary depending on the position of the sensors on the neck. This would help in understanding the feasibility of a continual signal from a collar-mounted ECG in freely moving animals, which would inevitably move around the neck. To investigate this, a trial was conducted on an awake, freely moving dog (male, German shepherd). In this trial, the 3 sensors were kept in the same alignment (green, black, and red) and the same distance apart (45°) but were moved about the neck to hit all compass points (see

Figure 23), 8 different positions in total. For example, 1 of 8 sensor positions would be green at N, black at NE and red at E.

Figure 23: Image to illustrate the compass point positioning of the ECG sensors during trials to assess how signal quality differed with sensor positions about the neck.

Attempts were made to trial different sensor locations in 3 macaques under GA. However, due to time restraints and positioning of the animals, these trials were unsuccessful at detecting an R wave.

6.3.4 Comparison to a standard ECG

To confirm the accuracy of the heart rate detected by the neck-based ECG, a comparison to a standard ECG device was conducted. For this, both the neck-based ECG and a veterinary clinical ECG device (Cardipia200 – Type 203RV, TRISMED Co Ltd) were used simultaneously. The neck ECG was set up as previously using the limb plate electrodes, and the standard device was connected to the animal via 4 ECG cables and crocodile clips (yellow clip on the left foreleg, red on the right foreleg, green on the left hindleg, and black on the right hindleg). The standard ECG was set-up by a veterinarian, Dr Caroline Bergmann, with years of experience in ECG set up. The internal clocks for each ECG device were time-matched and an interference event (simultaneous finger tap on the sensors) was used as a reference point within the signals to align them.

Attempts to obtain a sufficient signal for comparison were unsuccessful in 3 macaques under GA (see section 2.4 for GA protocols). We were unable to obtain a clear enough signal from either the standard or neck-based ECG device during the time available for data collection. This is

partly a result of restricted access and a time limit to obtain a signal, to avoid disruption to the neuroscience research. It was also partly caused by some failings with the neck-based ECG caused by degradation of the sensors, which will be discussed later in section 6.4.5.

A collaboration with a veterinary practice (Hilltop Veterinary Centre) and a veterinarian (Dr Hannes Bergmann) enabled this comparison study to be continued using dogs undergoing CT scans for veterinary diagnostic purposes. Propofol was used for induction of GA and maintained using volatile anaesthesia (isoflurane). Time and access to the dogs was still restricted, to avoid disruption to the scan, but was more permissive than with the macaques. Comparison trials were attempted on 3 female dogs: a German shepherd, a Labrador, and an English Cocker Spaniel.

6.3.5 Trials on a new species: Sheep

To determine whether the neck-based ECG could detect R waves in more species, a trial was conducted on a female sheep, under GA, at the University of Nottingham. The sheep was undergoing a surgical procedure as part of scientific protocols under the appropriate PPL authority as required by A(SP)A (Animals (Scientific Procedures) Act, 1986). The sheep was sedated with ketamine and midazolam and then a cannula was placed, followed by induction with alfaxalone. GA was maintained using volatile anaesthesia (isoflurane). Standard monitoring of vital signs was conducted, as described in macaques in section 2.4.6. The neck of the sheep was shaved as part of the procedure. Attempts were made to get an ECG trace with both the silicon and limb plate electrodes at various positions around the neck.

6.4 Results

6.4.1 Detecting an ECG trace from the neck

A successful ECG trial (i.e. the presence of R waves) was achieved in 5 out of 8 male macaques under GA and 2 out of 2 macaques under restraint.

Figure 24 contains segments from the 5 successful ECG traces obtained from macaques under GA. Limb plate electrodes were used on macaque WE and the silicon sensors for the other 4 macaques. Macaques were in a side-lying position during collection. With the exception of UC, sensors were placed on the side facing upwards, 45-90° apart with the central sensor sitting approximately under the ear. In macaque UC, sensors were placed around the back of the neck, 45-90° apart with the central sensor sitting in the middle of the nape. R waves are clearly visible in all traces, but R wave amplitudes varied, ranging from a mean of 3.0mv (± 0.2 SD) in macaque UT to 0.6mv (±0.09 SD) in UC.

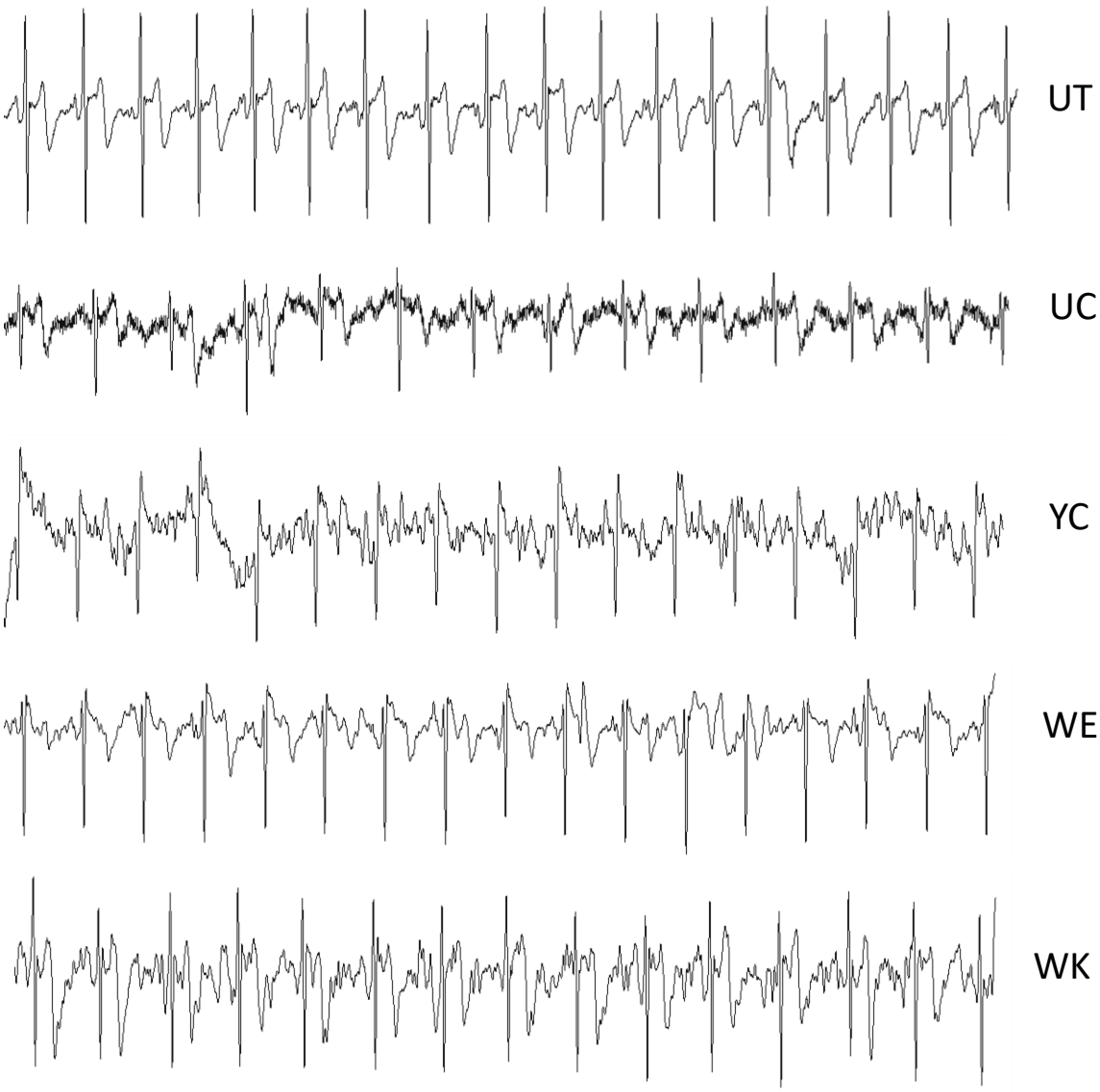

Figure 24: Traces obtained using the neck-based ECG from 5 male macaques (UT, UC, YC, WE, WK) under GA.

Figure 25 shows examples of the traces obtained when using the neck-based ECG device on macaques whilst they were restrained in a primate chair. Limb plate electrodes were used on animal PG and silicon sensors on animal WN. Sensors were placed at the back of the neck, 45-90° apart with the central sensor being in the middle of the nape. In both cases, R waves are clearly visible, but the magnitude of the R waves is much bigger in WN (mean 3.5mv ± 0.2 SD) compared to PG (mean 1.1mv ± 0.1 SD).

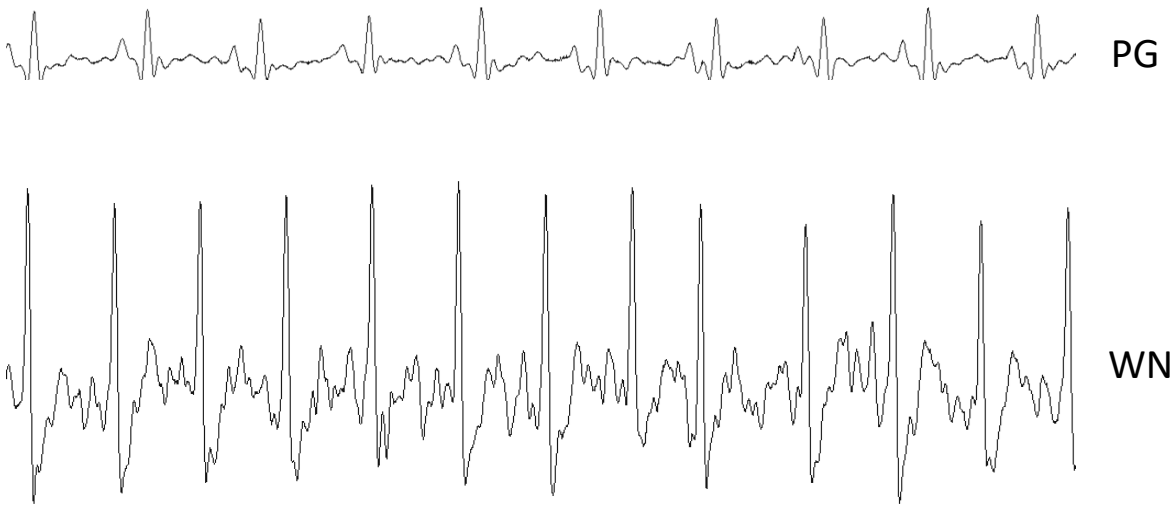

Figure 25: Traces obtained from the neck-based ECG from 2 male macaques (PG, WN) whilst awake and restrained in a primate chair.

6.4.2 Positioning of the sensors

The traces in Figure 26 represent the largest (a) and smallest (b) R-waves seen when positioning the ECG at 8 different locations around the neck of an awake German shepherd. The R-wave amplitudes in these traces average (a) 1.8mv (±0.2 SD) and (b) 0.9mv (±0.2 SD). In all 8 locations, an ECG signal could be detected using the pronged sensors and electrode gel as a conductive medium.

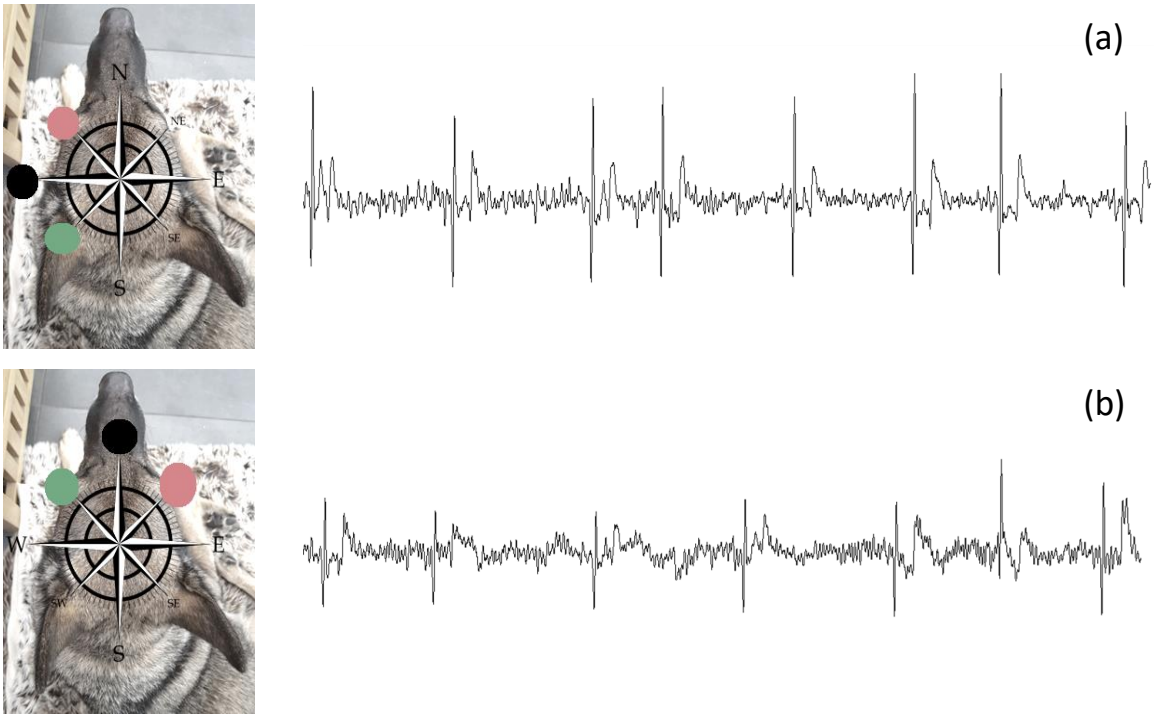

Figure 26: ECG traces from awake German shepherd dog and a picture diagram of sensor placement. Trace a) was obtained when sensors were placed in positions SW (green), W (black), and NW (red). Trace b) was obtained when sensors were placed in positions NW (green), N (black), and NE (red).

6.4.3 Comparison to a standard ECG

Due to encountering issues with obtaining signals and problems with sensors (discussed in section 6.4.5), we were only able to successfully obtain a comparison dataset from 1 out of 3 dogs (Cocker Spaniel). Figure 27 shows an image of the alignment of the paper printouts of the ECG signals obtained from the neck-based (top) and the standard (bottom) ECG device. Traces were printed to the same scale (25mm per second) and could be aligned precisely due to simultaneous occurrences of noise in the trace due to the interference event. R waves line up well and the RR intervals appear to match, with 13-14 squares (1mm) between each R wave. This equates to RR intervals of 520-560 ms. Using the VitalStore software, RR intervals can be measured digitally and more precisely. This was done on the same ECG trace and the results can be seen in Figure 28.

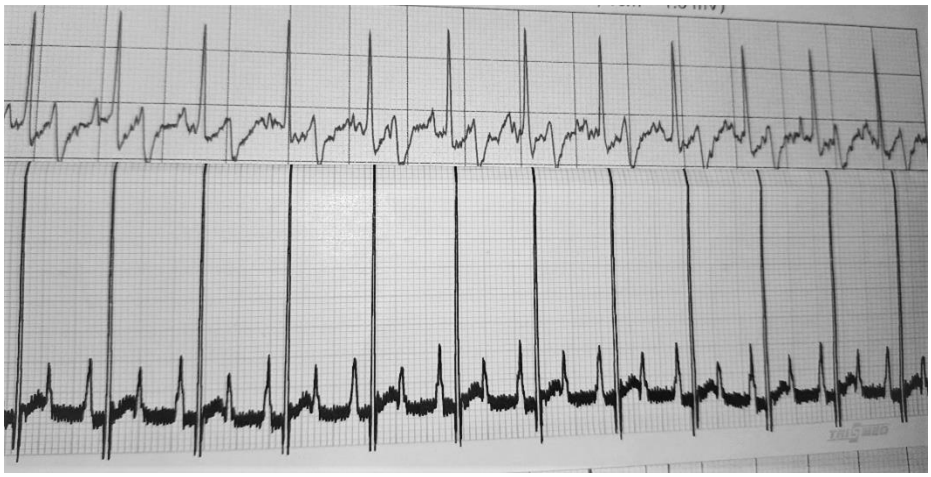

Figure 27: Image of paper ECG traces from neck-based ECG (top) and standard ECG (bottom) taken simultaneously from a Cocker Spaniel dog. The R waves from each trace have been aligned using an interference event. Each trace is printed to a scale of 25mm per second, with each small square being 1mm.

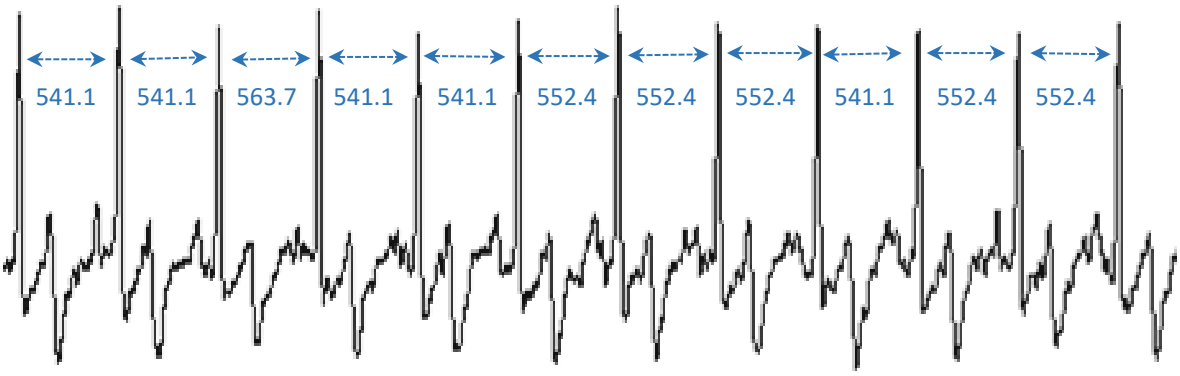

Figure 28: The same trace from the neck-based ECG (Cocker Spaniel) as seen at the top of Figure 27. This trace is digitally copied from the VitalStore software and RR intervals have been calculated between all R waves and shown underneath the arrows that span each interval.

6.4.4 Trials on a new species: Sheep

R waves were successfully detected when the limb plate electrodes were placed in positions N, NE, and E (using the same compass as that seen in Figure 23). A portion of the ECG trace obtained in this position can be seen in Figure 29. During this trial, there were a lot of times when R waves could not be detected and there were multiple experiences of 'signal dropout,' which is explained in section 6.4.5.

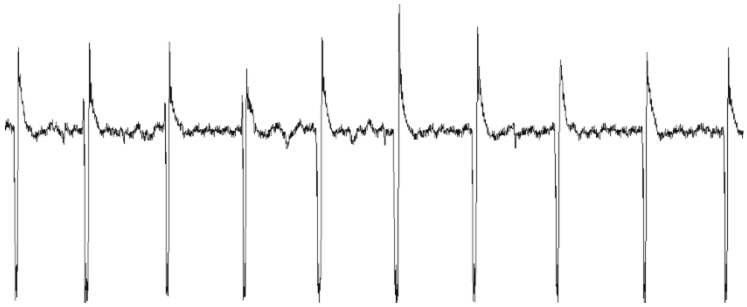

Figure 29: Trace from the neck-based ECG from a female sheep.

6.4.5 Issues that arose during testing

At various points in the trials using the neck-based ECG, there were a few reoccurring problems. These problems have either been resolved or have a plan to be addressed in further developments.

Firstly, there were several instances of 'signal dropout.' This resulted in the dropout of any ECG trace and the appearance of a flat line on the monitor. Attempts to resolve this were not always successful, but adjustment or resetting of sensors would sometimes restore the ECG signal.

Secondly, there were occasions where it just did not seem possible to detect R waves. This occasionally persisted following efforts to improve contact by repositioning sensors or adding more conductive medium.

Both of these problems were suspected to be from poor signal contact. The dropout occurred due to saturation of the input amplifiers resulting in high DC voltage (Keith Simpson, personal communication, November 2021). This occurs because of extensive skin or electrode resistance and only happened when using the silicon sensors. The frequent application of gel to the sensors to form a conductive medium caused degradation of the connections and increased the electrode resistance. Attempts were made to resolve this through cleaning but in the end the connectors for the sensors needed to be replaced. This did seem to stop the dropout at first but there was one instance of dropout during our final trial with the sheep.

Finally, when attempting to clean the silicon sensors using alcohol, the silicon brushes disintegrated and fell off. Fortunately, the sensors were able to be replaced quickly by the industrial partner but clearly caused disruption to the trial planned for that day.

6.5 Discussion

The success of these trials is evidence that a clear ECG signal, with distinguishable R waves, can be detected from the neck of multiple species. RR intervals are comparable between a standard ECG device and the neck-based ECG, with traces from both devices able to be aligned by their R-waves. This makes the neck-based ECG device a viable, non-invasive tool for heart rate monitoring in animals. The ability to measure RR intervals allows for the assessment of HRV, which can offer a valuable contribution to the understanding of welfare. Development of this ECG is still ongoing, and adaptations need to be made to improve the reliability of the device. More work is required to determine whether this device could be used in freely moving animals. R waves were able to be detected in all positions around the neck, although with varying magnitude, implying that movement of the collar in a freely moving animal should not affect the reliability of the device.

In its current state the ECG device often still requires the use of conductive medium between the sensors and the skin to increase contact. Furthermore, the specialised prongs

designed for this study did not appear to be effective at penetrating through the hair to get better contact with the skin. In most, but not all ECG trials, signal detection was more successful when using the metal, limb plate electrodes rather than the pronged sensors.

Although the addition of special-purpose, low noise amplifiers were used to amplify the signal whilst reducing noise, this does not filter noise from muscle activation and movement. This presents a problem for application in freely moving animals and further development would be required to minimise this noise.

6.5.1 Further development

The specialised, conductive prongs can be improved upon. Originally, a rubber material was chosen for the pronged sensors as it was envisaged that this would make them flexible, aiding in hair penetration. In practice, the rubber prongs are quite rigid. Furthermore, the connectivity with the connecting wires was quite poor, resulting in increased resistance and reduced signal quality with time. Therefore, trying an alternative sensor would be beneficial. Whilst still pronged, the sensor could be made from stainless steel with slightly broader and shorter prongs. This would hopefully improve hair penetration, allowing for better contact with the skin and reducing the need to use gel as a conductive medium. Furthermore, it would also improve connectivity to the wires. This design has already been discussed with the industrial collaborator and would be the next stage of development.

Head and neck movements are extremely common behaviours, so the ECG signal is likely to be frequently interrupted by these factors in mobile animals. To adapt the ECG device for use on freely moving animals, greater filtering can be further applied (either hardware or firmware) to attempt to obtain a more consistent, reliable signal for measure of HRV. Digital signal processing techniques can be applied to extract the ECG from a wandering baseline signal, provided that the movement signal does not saturate the amplifier. Further filtration of the signal may change the morphology of the QRS complex. Whilst this may make the obtained signal unsuitable for some

clinical observations, as long as the R peaks are not lost, heart rate and R-R intervals can still be measured, which is the main objective for this study.

EMG noise presents a more difficult problem. ECG and EMG signals are very similar, with the PQRST complex significantly overlapped by the spectral content of muscle activity (Kher, 2019). Therefore, an EMG signal may interfere or completely obscure the ECG signal. The current concept has no allowance or method to remove these signals. Although it has not yet been trialled, it is possible that some extraction techniques may be able to target the unique QRS complex within the mixed EMG and ECG signal, allowing HRV measurements to be made. It is also possible that collar design could help to minimise noise. A wider collar would be less subject to movement than a narrower collar, reducing the subsequent artefacts.

Currently, there is a small enclosure that houses the ECG system (Sapphire One transmitter) which is connected to the sensors in the collar via a standard ECG cable. The overall aim is to have the sensors and amplifier all housed within the collar itself. This should reduce the tendency to pick up additional noise and improve performance overall. Once this has been achieved, an appropriate collar can be designed to house the ECG system. An ECG receiver would still need to be within close proximity to be able to detect and record the signal. The vision for collar design would be similar to that used for accelerometers in previous chapters (as described in section 2.1), with adaptations where required. If the ECG device can be enclosed within a collar, then it will be able to be fitted for trials on freely moving animals.

6.5.2 Future potential

Having successfully detected R waves from the neck ECG in humans, dogs, macaques, and sheep, this ECG device clearly has the potential for applications across many different species. Previously, animal studies requiring an ECG involved more invasive methods, such as GA or implantable telemetry device, which could confound results or compromise welfare. Non-invasive chest straps have been a popular tool for measuring HRV in many species, particularly farm

animals (von Borell et al., 2007). Whilst these chest straps still offer great value, a neck-based ECG may offer further benefits to be considered. For example, the neck may be more accessible in some species compared to the torso, allowing for a quicker and simpler fit.

Once the ECG device can be tried on moving animals, then the level of interference caused by movement and muscle activation will become clear. The unique structure of the QRS complex means that it may still be detectable amongst EMG noise. However, the obscuring of ECG signals during movement still presents in the popular ECG chest straps (Jonckheer-Sheehy et al., 2012) and so without trials, we cannot predict if this is achievable. The design of this ECG aims to be minimally intrusive, such that the device could be worn continually over several days and data collected remotely. Therefore, even if movement obscures the R wave completely, HRV should still be measurable in the moments when the animal is stationary. Thus, HRV should be obtainable during periods of rest, and this would offer the opportunity to measure HRV overnight. Studies conducted on macaques reported using HRV at night as a measure of stress, as it represents the nadir of activity for diurnal animals (Shively et al., 2020; Silverstein-Metzler et al., 2022). In captive animals, this represents a period with the least disruption and so is relatively independent of extrinsic factors, such as locomotion and human presence.

This continual, remote method of ECG detection offers further benefits to HRV monitoring too. Longer monitoring periods gives the opportunity to gain sufficient data for all measures of HRV. Often HRV studies use just 5 minutes (short term) or less (ultra-short term) of ECG data to calculate their HRV variables. Twenty-four-hour HRV monitoring provides the most valid measure of many time and frequency domains (Shaffer & Ginsberg, 2017). Furthermore, in the context of research macaques, continuous monitoring allows for the capture of spontaneous events that would otherwise be missed. For example, HRV could quantify the social stress generated by a fight with a conspecific. The stress response to social challenges is key to obtaining a picture of overall welfare, particularly in social species such as macaques. When measuring HRV in rats,

greater autonomic imbalance was reported following a social challenge compared to non-social stressors, including restraint, swimming, and a shock-probe test (Sgoifo et al., 1999).

ECG devices are typically used as a clinical tool for diagnosis. This device could also aid in the clinical monitoring of animals, dependent on the quality of the signal. In the context of macaques used in neuroscience research, post-operative recovery presents challenges within the ANS which require regular monitoring to establish intervention points. This device could be a tool for remote and continuous assessment of cardiac performance – a measure which is not usually obtained. R waves may be of immediate benefit, for example in identifying sinus arrhythmia (Richig & Sleeper, 2019), but a clear view of the QRS complex or P or T waves may be required to make further diagnosis. However, this ECG should be used in combination with standard cage-side assessment protocols and not used as a substitute.

6.5.3 Conclusion

HRV is a useful parameter to further the understanding of animal welfare. For the most accurate measure of HRV, an ECG device is required. By establishing that R waves can be detected from ECG sensors placed on the neck of animals, this study offers a non-invasive approach for remote, continuous recording of HRV in multiple species. Specifically, for NHPs, the neck placement of this device minimises the risk of animal interaction with the device and could be an incredibly valuable tool for quantifying the ANS responses to multiple experimental and husbandry related stressors.

Chapter 7 - Discussion

Whilst it is still considered vital to use animals as models in biomedical research, it is imperative that we continually seek out refinements to experimental and husbandry practices and strive for better welfare standards. As highly sentient species engaging in complex, long-term studies, the propensity to suffer is particularly high in non-human primates (NHP) used in neuroscience research (Perretta, 2009). Endeavouring to optimise welfare is both a moral and legal obligation and will result in better quality science (Prescott & Lidster, 2017). For example, in the context of neuroscience, research suggests that stress can influence areas of cognitive functioning (Shields et al., 2016) and so can affect performance in standard cognitive and behavioural tests.

To improve animal welfare, we must first be able to assess it. This is essential for effecting changes in legislation and policies, which lean on scientific and quantitative evidence in order to make their decisions on animal research practices (Mellor et al., 2009). In this thesis, I have presented methodologies for the assessment of the welfare of NHPs used in neuroscience research. Welfare parameters were selected for their non-invasive nature, allowing for welfare assessment to be conducted alongside neuroscience experiments with minimal disruption to their primary research purpose. The implementation of such measures aimed to explore the acute and long-term welfare impact of experimental protocols and husbandry regimes. For this purpose, I used accelerometers to evaluate physical activity and behaviour, faecal and hair samples to monitor cortisol response and a customised ECG device to assess heart rate variability (HRV).

7.1 General anaesthesia

An essential procedure for many neuroscience protocols is the administration of general anaesthesia (GA). Although used as a refinement, to optimise data acquisition and reduce awareness of potentially painful or stressful procedures (NC3Rs, 2021), GA is a probable stressor

that can impact animal welfare. In chapters 3 and 5 a decline in activity levels and an increase in faecal cortisol metabolites (FCM) can be seen in response to GA procedures. Both of these changes may be indicative of reduced welfare. In response to a GA alone (MRI), both the activity and FCM responses return to baseline levels within several days. Following a surgical procedure, a greater and more prolonged increase in FCM was observed and although activity levels returned to baseline 3 days earlier compared to MRI, this was followed by an increase in activity rather than the maintenance of baseline. This evidence suggests that whilst administration of GA may be an acute stressor, surgical procedures have a more profound and long-lasting impact on macaques. Similar differences were detected in the cortisol concentration in hair samples (HCC), as surgical procedures were predicted as drivers of increases in HCC but MRIs resulted in no obvious change to HCC (chapter 5). The welfare implications from each parameter (HCC, FGM, and activity levels) appear to be consistent in highlighting that MRI events are a source of acute stress but surgical procedures have a greater, more prolonged effect, potentially contributing to chronic stress. This result was not unexpected, as surgical protocols involve additional physiological stress to GA, such as tissue trauma and post-operative pain (Desborough, 2000). Furthermore, the nature of many surgical procedures imposes physical change to the animal which may influence their welfare state. For example, head-holding devices and recording chambers protrude upwards from the head. Furthermore, cranial implants can become infected (Johnston et al., 2016) and require regular cleaning of the wound margins. There is some evidence that the repetition of GA and the length of a GA influence the recovery of physical activity levels in macaques (chapter 3). Further understanding of how such parameters effect GA recovery could offer potential refinements. For example, would macaque welfare benefit from more frequent but shorter GA procedures or less, longer ones? Application of the methods presented in this study would be beneficial for determining these effects.

7.2 Wearable sensors

Using accelerometers offers an opportunity for remote monitoring of macaque behaviour. Actigraphy data at the most basic level offers value in monitoring changes in physical activity (chapter 3) but we have also seen how actigraphy data can be used to predict behavioural profiles (chapter 4). With an overall accuracy of 69%, the rule-based model from chapter 4 offers an insight into the behavioural profile of rhesus macaques. Whilst daily behavioural assessments through observations are carried out as part of the standard practice in NHP facilities, the use of accelerometers offers an opportunity for a continuous assessment, which is unaffected by the presence of an observer. Of particular interest is the high detection rate (88%) of pacing behaviour from accelerometer data. Although the cause of pacing can be unclear, it is considered an abnormal behaviour indicative of suboptimal conditions in NHPs (Lutz et al., 2003) and monitoring of such behaviours plays a vital role in welfare assessment (Mason & Latham, 2004).

This study has provided proof of concept that an ECG signal can be obtained from around the neck (chapter 6). Measures of HRV obtained from this novel device are comparable to a standard ECG. Whilst there is still development required to optimise this device, it offers an exciting opportunity as a more refined, less intrusive tool for obtaining HRV measures from freely moving macaques. Furthermore, it has also been shown that this neck-based ECG device can be used in other species: dogs and sheep.

7.3 Individual variation

A high degree of inter-individual variation was observed in measures of cortisol and activity (chapter 3 and 5), as well as being observed between the behavioural profiles of macaques when creating the rule-based model for actigraphy data (chapter 4). This emphasises the importance of accounting for biological variation, not only in welfare studies but in the results obtained from neuroscience experiments too. This is particularly important in studies using NHPs, where the

animal numbers for each experiment are low (e.g. 2 or 3 animals). Biological variation is not inherently bad, as it is a reflection of what is true in nature, but it should be accounted for in the design of experiments and when interpreting results (Voelkl et al., 2020). Furthermore, consideration should be given to individual coping styles, suggesting that the same procedure in two animals may have a very different welfare impact (Koolhaas et al., 2010). Employing continuous methods of welfare assessment, such as those used in this study, can highlight individual variation, and allows for within-individual comparisons to be made. This may offer opportunity for refinements to be made on an individual basis.

7.4 Social stress

During analysis of individual profiles, HCC appeared to increase following periods of social stress (chapter 5). Unstable social groups within the restrictive living space available to laboratory animals can result in an increased incidence of agonistic behaviour and fighting. This can directly impact welfare, resulting in injury and possibly requiring separating or rearranging of social groups (Hannibal et al., 2017). Consideration of contingent harms forms an essential part of overall welfare assessment and should be considered when justifying animal research (Hubrecht, 2014). Furthermore, affecting refinements based on contingent harms is less likely to impact negatively on scientific output than direct changes to experimental protocols and thus, researchers may be more perceptive to change. For example, when social stress is unavoidable but predictable, further known stressors (e.g. surgeries) during this period could be avoided or attempts could be made to minimise social stress through compatibility testing prior to grouping animals.

Although single housing of macaques did not affect HCC in this study, it has been highlighted as a major psychological stressor for NHPs (DiVincenti & Wyatt, 2011; Rennie & Buchanan-Smith, 2006b) and this is reflected in the requirement for social housing, where possible, in the legislation governing animal research (Animals (Scientific Procedures) Act, 1986).

In the absence of compatible social partners, efforts should be made to offer alternative socialisation opportunities. The benefits of this are indicated in this study where indirect social contact, by housing macaques in the same room as others, instigated a change to the behavioural profile of a female macaque indicative of increased socialisation (see 'Case study 2' in chapter 4).

7.5 Using multiple measures of welfare

Welfare assessment of NHPs can be further improved through the combination of welfare parameters. Welfare measures may be contradictory for various reasons, such as individual variation and differences in coping styles between animals (Broom, 2014; Mason & Mendl, 1993). Therefore, to better gain the full picture of animal welfare, a range of measures is required. All parameters discussed in this thesis could be obtained simultaneously to give an overview of both physiological and behavioural changes. The ECG and accelerometers could feasibility be fitted inside the same collar, allowing for HRV and actigraphy collection at the same time without any further disruption to the animals. Furthermore, some welfare measures may serve to provide context to others. Cortisol and activity levels are correlated, with increased cortisol seen during periods of increased activity (exercise) (Otovic & Hutchinson, 2015). Similarly, activity increase also leads to increased cardiac activity and changes to HRV. The hypothalamic-pituitary-adrenal (HPA) axis and the autonomic nervous system (ANS) are highly coordinated and interconnected. When provided with information about both systems, the inter-relation between the two may aid in the understanding of perceived stress (Rotenberg & McGrath, 2016). Vagal activity, which can be determined through measures of HRV, is thought to inhibit HPA axis regulation. Combining measures of HRV and cortisol response could be enlightening about the mechanisms of the stress response. In one study, changes in HRV during the anticipation of a stressful situation related to cortisol reactivity following the stressor (Pulopulos et al., 2018). This is potentially valuable in monitoring individual differences in stress regulation. Using a combination of HRV and cortisol

measures is already being applied and aiding in our understanding of animal welfare (Bergamasco et al., 2010; Schmidt et al., 2010; Shively et al., 2020).

7.6 Cumulative experience

It is a legal requirement that assessment of an animal's cumulative effects is considered when using them as research models (Animals (Scientific Procedures) Act, 1986). Whilst individual experiences can have a positive or negative impact on welfare in the short term, what is crucial is how these effects balance to determine an animal's cumulative experience. Although qualitative data (e.g. the opinions of those working directly with animals) does play an important role in the understanding of animal experience, quantitative and objective measures of cumulative experience are required for a better understanding and to provide a basis for refinements (Animals in Science Committee, 2017; Pickard, 2013). Whilst measures such as telomere attrition (Bateson, 2016) and grey matter volume (Lee et al., 2009) are valuable for cumulative assessment, widespread uptake of such measures would be difficult to implement due to the complex analysis required. The ASC recommends that obtaining physiological and behavioural indicators of cumulative experience alongside ongoing regulated procedures could provide insight whilst avoiding additional harms (Animals in Science Committee, 2017). This also allows for consideration of the compound effect of multiple positive and negative experiences at once, which may better reflect the overall experience of laboratory animals. The nature of the parameters used in this study make them ideally suited to the assessment of cumulative experience. The non-invasive and continuous data collection protocols allow for longitudinal assessment, including repeated measures of the same stressors and consideration of contingent harms.

The cumulative effect of exposure to multiple MRIs, on activity levels (chapter 3), was unclear in this study but there was some indication that previous MRI exposure may influence a macaque's recovery from GA. Furthermore, in chapter 5 a possible cumulative effect was seen in

HCC when social stress occurred in close proximity to experimental stressors. However, there was no overall changes in HCC that would indicate cumulative suffering. Whilst the conclusions on cumulative experience in this study were not profound, there is still inherent value in the methodologies used. Combining longitudinal measures of acute and chronic stress can provide a life-time account of the stress experience of individual macaques. This could be hugely beneficial to the cumulative assessment of macaques used in neuroscience research.

7.7 Future benefits and limitations

The welfare parameters used in this study are not intended to replace existing cage-side assessments that are used to monitor the day-to-day wellbeing of NHPs. They are instead proposed as additional measures that may provide a more quantitative and objective measure. Whilst direct observations are limited to the routines of care staff and researchers, the continuous monitoring permitted by wearable sensors (accelerometers and ECG), allow for welfare monitoring throughout all times of the day and night. Measuring night-time activity in captive macaques could be illuminating to their wellbeing as a potential indicator of sleep (Cortes et al., 2016; Stanwicks et al., 2017). Actigraphy has been used extensively to study sleep in humans (Ancoli-Israel et al. 2003). Insight into the sleep of NHPs would contribute to the understanding of their welfare, as sleep quality correlates with well-being in humans. This is something that has already been trialled by a member of our lab using the accelerometer data obtained from this study and shows promise as a valuable welfare measure. Furthermore, the neck-based ECG device (chapter 6) was intentionally designed for continuous wear, including overnight. Assessing HRV during the night can be beneficial, as ECG traces are less likely to be obscured by macaque movement and potential stressors (e.g. human presence) are minimised (Shively et al., 2020; Silverstein-Metzler et al., 2022).

One of the key issues with implementing the 3Rs to NHP research is a reluctance to change. Due to the strict legal and financial limitations on primate use and scientists being subjected to

enormous pressure to publish in high-impact journals, developing and validating refinements cannot always be a priority (Lemon, 2018; Prescott et al., 2017). Furthermore, NHP studies are usually small (i.e. use small numbers of NHPs) and so any attempts to change that ultimately have a detrimental effect can profoundly influence the outcomes of the study. The methods applied in this study could help to find and implement refinements despite these issues. Welfare assessment alongside ongoing research will not only provide further evidence to quantify the impact of experimental procedures but actively engaging neuroscientists with the welfare assessment of their own animals may provide a greater appreciation of the animals experiences. The opportunity for continuous, quantitative welfare monitoring may be beneficial in highlighting areas for improvement, in both husbandry and experimental practices, which would require minimal change to achieve. For example, case study 1 presented in chapter 4 in which altering the time of day a macaque was tested could lead to a more normal behavioural profile.

As with all measures of welfare, there are obstacles to overcome and limitations to consider. Although context can help with our understanding, physiological measures are only really an indication of arousal and do not point to the valence of affective states (Dawkins, 2006). For example, an increase in cortisol can be associated with both stress and excitement. Furthermore, there are likely to be incidences when certain welfare parameters are not appropriate. In the case of research animals, part of their scientific practices may involve the administration of drugs, which limits the value of some physiological measures. In this study, we saw how treatment with steroids (prednisolone and dexamethasone) caused a reduction in the cortisol detected in hair samples (chapter 5). Although measures in this study are defined as non-invasive, it is important to distinguish that there may still be some impact on the animals. Macaques may experience transient stress during hair sample collection, as clippers used to trim the hair produce noise and although will not hurt, may feel strange when touched to the neck. Deployment of collars did not appear to cause obvious distress to macaques, especially after habituating to them (usually took <1 day, with less time required after subsequent fittings).

However, NHPs are non-domesticated animals and so there is still a possibility that the collar could be poorly tolerated. Care should be taken to reduce any impact, such as fitting collars or collecting samples during pre-planned restraint (i.e. that taking place for neuroscientific experiments). Furthermore, careful monitoring should take place after collar fittings, particularly in animals that have not previously worn a collar, and collar fit should be appropriately adjusted to minimise discomfort.

The 'piggy backing' methodologies employed in this study did come with logistical challenges. Limited control over the specifics of the data obtained can prove challenging when aiming to test specific hypotheses. For example, to assess the impact of GA duration, this style of research is reliant on the occurrence of different length GAs as part of the neuroscience protocols. Furthermore, busy schedules and different priorities can sometimes make collaboration between welfare research and neuroscience challenging. In order to continue these types of research collaborations, sufficient financial and personnel resources are essential.

7.8 Conclusion

In summary, in this thesis I have explored the welfare assessment of rhesus macaques used in neuroscience research. The methods and parameters used permit longitudinal and repeated assessment of welfare throughout the lifetime of individual macaques. This thesis has explored physiological and behavioural responses to specific, transient neuroscience stressors and also considered the effects of contingent harms and cumulative experience. Advancing our understanding of macaque welfare through quantitative and objective assessment is crucial to our perception of their quality of life. Continuous assessment is essential for identifying refinements and striving to improve macaque welfare. Not only is this an ethical and moral consideration, but it is also a legal obligation for anyone using macaques as research models and engaging in better welfare assessment honours their commitment to the 3Rs.

Bibliography

Accorsi, P. A., Carloni, E., Valsecchi, P., Viggiani, R., Gamberoni, M., Tamanini, C., & Seren, E. (2008). Cortisol determination in hair and faeces from domestic cats and dogs. *Gen Comp Endocrinol, 155*(2), 398-402. doi:10.1016/j.ygcen.2007.07.002

Anestis, S. F. (2009). Urinary cortisol responses to unusual events in captive chimpanzees (Pan troglodytes). *Stress-the International Journal on the Biology of Stress, 12*(1), 49-57. doi:10.1080/10253890802041308

Animal Welfare (Sentience) Act. (2021). Retrieved from https://publications.parliament.uk/pa/bills/lbill/58-02/004/5802004_en_2.html#pb1-l1g1

Animal Welfare Act. (2006). Retrieved from https://www.legislation.gov.uk/ukpga/2006/45

Animals (Scientific Procedures) Act. (1986). *(Amendment Regulations 2012).* H.M.S.O, London. Retrieved from https://www.gov.uk/guidance/research-and-testing-using-animals

Animals in Science Committee. (2017). *Review of harm-benefit analysis in the use of animals in research*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/675002/Review_of_harm_benefit_analysis_in_use_of_animals_18Jan18.pdf

Ash, H., Smith, T. E., Knight, S., & Buchanan-Smith, H. M. (2018). Measuring physiological stress in the common marmoset (Callithrix jacchus): Validation of a salivary cortisol collection and assay technique. *Physiology & Behavior, 185*, 14-22. doi:10.1016/j.physbeh.2017.12.018

Azevedo, A., Bailey, L., Bandeira, V., Dehnhard, M., Fonseca, C., de Sousa, L., & Jewgenow, K. (2019). Age, sex and storage time influence hair cortisol levels in a wild mammal population. *Plos One, 14*(8). doi:10.1371/journal.pone.0221124

Bahr, N. I., Palme, R., Mohle, U., Hodges, J. K., & Heistermann, M. (2000). Comparative aspects of the metabolism and excretion of cortisol in three individual nonhuman primates. *General and Comparative Endocrinology, 117*(3), 427-438. doi:DOI 10.1006/gcen.1999.7431

Barnett, J. L., & Hemsworth, P. H. (1990). The Validity of Physiological and Behavioral Measures of Animal-Welfare. *Applied Animal Behaviour Science, 25*(1-2), 177-187. doi:10.1016/0168-1591(90)90079-S

Barrasso, R., Bonerba, E., Ceci, E., Roma, R., Alo, A., Mottola, A., . . . Bozzo, G. (2020). Evaluation of the animal welfare during religious slaughtering. *Italian Journal of Food Safety, 9*(1), 39-43. doi:10.4081/ijfs.2020.8387

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:doi:10.18637/jss.v067.i01

Bateson, M. (2016). Cumulative stress in research animals: Telomere attrition as a biomarker in a welfare context? *Bioessays, 38*(2), 201-212. doi:10.1002/bies.201500127

Bateson, M., & Poirier, C. (2019). Can biomarkers of biological age be used to assess cumulative lifetime experience? *Animal Welfare, 28*(1), 41-56. doi:10.7120/09627286.28.1.041

Bateson, P. (1986). When to experiment on animals. *New Sci, 109*(1496), 30-32.

Bateson, P. (2004). Do animals suffer like us? the assessment of animal welfare. *Veterinary Journal, 168*(2), 110-111. doi:10.1016/j.tvjl.2003.12.002

Bateson, P. (2005). Ethics and behavioural biology. *Advances in the study of Behavior, 35*, 211-233.

Bateson, P., Johansen-Berg, H., Jones, D. K., Keverne, E. B., Matthews, P. M., Milner, A. D., . . . Peck, H. (2011). Review of research using non-human primates. . *Report chaired by Professor Sir Patrick Bateson FRS, London, BBSRC/MRC/NC3Rs/Wellcome Trust*. Retrieved from https://mrc.ukri.org/documents/pdf/bateson-review-of-non-human-primates/

Bechshoft, T. O., Riget, F. F., Sonne, C., Letcher, R. J., Muir, D. C., Novak, M. A., . . . Dietz, R. (2012). Measuring environmental stress in East Greenland polar bears, 1892-1927 and 1988-

2009: what does hair cortisol tell us? *Environ Int, 45*, 15-21. doi:10.1016/j.envint.2012.04.005

Bentham, J. (1823). *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Bergamasco, L., Osella, M. C., Savarino, P., Larosa, G., Ozella, L., Manassero, M., . . . Re, G. (2010). Heart rate variability and saliva cortisol assessment in shelter dog: Human-animal interaction effects. *Applied Animal Behaviour Science, 125*(1-2), 56-68. doi:10.1016/j.applanim.2010.03.002

Berman, M., Gaillard, J. M., & Weimerskirch, H. (2009). Contrasted patterns of age-specific reproduction in long-lived seabirds. *Proc Biol Sci, 276*(1655), 375-382. doi:10.1098/rspb.2008.0925

Bernacky, B. J., Gibson, S. V., Keeling, M. E., & Abee, C. R. (2002). Nonhuman primates. In J. G. Fox, L. C. Anderson, F. M. Loew, & F. W. Quimby (Eds.), *Laboratory Animal Medicine: 2nd Edition* (2 ed.). Elsevier Science (USA): Academic Press. (Reprinted from: 2nd Edition).

Bethell, E. J. (2015). A "How-To" Guide for Designing Judgment Bias Studies to Assess Captive Animal Welfare. *Journal of Applied Animal Welfare Science, 18*, S18-S42. doi:10.1080/10888705.2015.1075833

Bisgaard, T., Klarskov, B., Kehlet, H., & Rosenberg, J. (2002). Recovery after uncomplicated laparoscopic cholecystectomy. *Surgery, 132*(5), 817-825. doi:10.1067/msy.2002.127682

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., . . . Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiol Behav, 92*(3), 375-397. doi:10.1016/j.physbeh.2007.02.003

Bourque, S. L., Adams, M. A., Nakatsu, K., & Winterborn, A. (2010). Comparison of buprenorphine and meloxicam for postsurgical analgesia in rats: effects on body weight, locomotor activity, and hemodynamic parameters. *J Am Assoc Lab Anim Sci, 49*(5), 617-622.

Boyce, W. T., Champoux, M., Suomi, S. J., & Gunnar, M. R. (1995). Salivary cortisol in nursery-reared rhesus monkeys: Reactivity to peer interactions and altered circadian activity. *Developmental Psychobiology, 28*(5), 257-267. doi:10.1002/dev.420280502

Bracke, M. B. M., & Hopster, H. (2006). Assessing the importance of natural behavior for animal welfare. *Journal of Agricultural & Environmental Ethics, 19*(1), 77-89. doi:10.1007/s10806-005-4493-7

Brambell, F. W. R. (1965). *Report of the Technical Committee to enquire into the Welfare of animals kept under intensive livestock husbandry systems*. London, UK

Brent, L. J. N., Semple, S., Dubuc, C., Heistermann, M., & MacLarnon, A. (2010). Social Capital and Physiological Stress Levels in Adult Female Rhesus Macaques (Macaca Mulatta). *American Journal of Primatology, 72*, 68-68. Retrieved from <Go to ISI>://WOS:000280720400135

Broom, D. M. (1988). The Scientific Assessment of Animal-Welfare. *Applied Animal Behaviour Science, 20*(1-2), 5-19. doi:10.1016/0168-1591(88)90122-0

Broom, D. M. (2014). *Sentience and Animal Welfare*. Wallingford, Boston: CABI.

Broom, D. M., & Johnson, K. G. (1993). *Stress and animal welfare* (1st ed.). Dordrecht, Boston, London: Kluwer Academic Publishers.

Brown, D. D., Kays, R., Wikelski, M., Wilson, R., & Klimley, A. P. (2013). Observing the unwatchable through acceleration logging of animal behaviour. *Animal Biotelemetry, 1*(20), 1-16.

Burnham, K. P., & Andersen, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach.* New York: Springer-Verlag.

Camus, S. M., Rochais, C., Blois-Heulin, C., Li, Q., Hausberger, M., & Bezard, E. (2014). Depressive-like behavioral profiles in captive-bred single- and socially-housed rhesus and cynomolgus macaques: a species comparison. *Front Behav Neurosci, 8*, 47. doi:10.3389/fnbeh.2014.00047

Capitanio, J. P., & Emborg, M. E. (2008). Contributions of non-human primates to neuroscience research. *Lancet, 371*(9618), 1126-1135. doi:Doi 10.1016/S0140-6736(08)60489-4

Capitanio, J. P., Mendoza, S. P., Lerche, N. W., & Mason, W. A. (1998). Social stress results in altered glucocorticoid regulation and shorter survival in simian acquired immune deficiency syndrome. *Proc Natl Acad Sci U S A, 95*(8), 4714-4719. doi:10.1073/pnas.95.8.4714

Carlitz, E. H. D., Kirschbaum, C., Miller, R., Rukundu, J., Mukungu, T., & van Schaik, C. P. (2015). New Insights into Hair Cortisol Analysis for Long-Term Stress Monitoring in Chimpanzees (Pan troglodytes). *Folia Primatologica, 86*(4), 257-258.

Carlitz, E. H. D., Kirschbaum, C., Stalder, T., & van Schaik, C. P. (2014). Hair as a long-term retrospective cortisol calendar in orang-utans (Pongo spp.): New perspectives for stress monitoring in captive management and conservation. *General and Comparative Endocrinology, 195*, 151-156. doi:10.1016/j.ygcen.2013.11.002

Carlitz, E. H. D., Miller, R., Kirschbaum, C., Gao, W., Hänni, D. C., & Van Schaik, C. P. (2016). Measuring Hair Cortisol Concentrations to Assess the Effect of Anthropogenic Impacts on Wild Chimpanzees (Pan troglodytes). *Plos One, 11*(4), e0151870. doi:10.1371/journal.pone.0151870

Carlsson, H. E., Schapiro, S. J., Farah, I., & Hau, J. (2004). Use of primates in research: A global overview. *American Journal of Primatology, 63*(4), 225-237. doi:10.1002/ajp.20054

Casal, N., Manteca, X., Pena, R., Bassols, A., & Fabrega, E. (2017). Analysis of cortisol in hair samples as an indicator of stress in pigs. *Journal of Veterinary Behavior-Clinical Applications and Research, 19*, 1-6. doi:10.1016/j.jveb.2017.01.002

Casey-Trott, T. M., & Widowski, T. M. (2018). Validation of an accelerometer to quantify inactivity in laying hens with or without keel-bone fractures. *Animal Welfare, 27*(2), 103-114. doi:10.7120/09627286.27.2.103

Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics, 4*(4). doi:10.15406/ijbsbe.2018.04.00125

Catipovic-Veselica, K., Amidzic, V., Durijancek, J., Kozmar, D., Sram, M., Glavas, B., & Catipovic, B. (1999). Association of heart rate and heart-rate variability with scores on the Emotion Profile Index in patients with acute coronary heart disease. *Psychological Reports, 84*(2), 433-442. doi:Doi 10.2466/Pr0.84.2.433-442

Chapa, J. M., Maschat, K., Iwersen, M., Baumgartner, J., & Drillich, M. (2020). Accelerometer systems as tools for health and welfare assessment in cattle and pigs - A review. *Behavioural Processes, 181*. doi:10.1016/j.beproc.2020.104262

Chapinal, N., de Passille, A. M., Pastell, M., Hanninen, L., Munksgaard, L., & Rushen, J. (2011). Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle. *Journal of Dairy Science, 94*(6), 2895-2901. doi:10.3168/jds.2010-3882

Chichagova, V., Georgiou, M., Carter, M., Dorgau, B., Hilgen, G., Collin, J., . . . Lako, M. (2023). Incorporating microglia-like cells in human induced pluripotent stem cell-derived retinal organoids. *Journal of cellular and molecular medicine, 27*(3), 435-445. doi:10.1111/jcmm.17670

Cloutier Barbour, C., Danforth, M. D., Murphy, H., Sleeper, M. M., & Kutinsky, I. (2020). Monitoring great ape heart health through innovative electrocardiogram technology: Training methodologies and welfare implications. *Zoo Biology, 39*(6), 443-447. doi:10.1002/zoo.21567

Cockrem, J. F. (2013). Individual variation in glucocorticoid stress responses in animals. *General and Comparative Endocrinology, 181*, 45-58. doi:10.1016/j.ygcen.2012.11.025

Colpaert, F. C., Dewitte, P., Maroli, A. N., Awouters, F., Niemegeers, C. J. E., & Janssen, P. A. J. (1980). Self-Administration of the Analgesic Suprofen in Arthritic Rats - Evidence of Mycobacterium-Butyricum-Induced Arthritis as an Experimental-Model of Chronic Pain. *Life Sciences, 27*(11), 921-928. doi:Doi 10.1016/0024-3205(80)90101-0

Cortes, J. A., Gomez, G., Ehnerd, C., Gurnsey, K., Nicolazzo, J., Bradberry, C. W., & Jedema, H. P. (2016). Altered activity-based sleep measures in rhesus monkeys following cocaine self-administration and abstinence. *Drug Alcohol Depend, 163*, 202-208. doi:10.1016/j.drugalcdep.2016.04.014

Cross, N., & Rogers, L. J. (2004). Diurnal cycle in salivary cortisol levels in common marmosets. *Developmental Psychobiology, 45*(3), 134-139. doi:10.1002/dev.20023

Culp, W. T., Mayhew, P. D., & Brown, D. C. (2009). The effect of laparoscopic versus open ovariectomy on postsurgical activity in small dogs. *Vet Surg, 38*(7), 811-817. doi:10.1111/j.1532-950X.2009.00572.x

Cusack, B., & Buggy, D. J. (2020). Anaesthesia, analgesia, and the surgical stress response. *Bja Education, 20*(9), 321-328. doi:10.1016/j.bjae.2020.04.006

Czoty, P. W., Gould, R. W., & Nader, M. A. (2009). Relationship Between Social Rank and Cortisol and Testosterone Concentrations in Male Cynomolgus Monkeys (Macaca fascicularis). *Journal of Neuroendocrinology, 21*(1), 68-76. doi:10.1111/j.1365-2826.2008.01800.x

d'Angelo, D., d'Ingeo, S., Ciani, F., Visone, M., Sacchettino, L., Avallone, L., & Quaranta, A. (2021). Cortisol Levels of Shelter Dogs in Animal Assisted Interventions in a Prison: An Exploratory Study. *Animals, 11*(2). doi:10.3390/ani11020345

Danbury, T. C., Weeks, C. A., Waterman-Pearson, A. E., Kestin, S. C., & Chambers, J. P. (2000). Self-selection of the analgesic drug carprofen by lame broiler chickens. *Veterinary Record, 146*(11), 307-311. doi:10.1136/vr.146.11.307

Darwin, C. (1873). *The expression of the emotions in man and animals*. New York,: D. Appleton and company.

Davenport, M. D., Tiefenbacher, S., Lutz, C. K., Novak, M. A., & Meyer, J. S. (2006). Analysis of endogenous cortisol concentrations in the hair of rhesus macaques. *Gen Comp Endocrinol, 147*(3), 255-261. doi:10.1016/j.ygcen.2006.01.005

Davies, J. R., Purawijaya, D. A., Bartlett, J. M., & Robinson, E. S. J. (2022). Impact of Refinements to Handling and Restraint Methods in Mice. *Animals : an open access journal from MDPI, 12*(17), 2173. doi:10.3390/ani12172173. (Accession No. 36077894)

Dawkins, M. S. (2003). Behaviour as a tool in the assessment of animal welfare. *Zoology, 106*(4), 383-387. doi:Doi 10.1078/0944-2006-00122

Dawkins, M. S. (2006). A user's guide to animal welfare science. *Trends in Ecology & Evolution, 21*(2), 77-82. doi:10.1016/j.tree.2005.10.017

Dawkins, M. S. (2008). The Science of Animal Suffering. *Ethology, 114*(10), 937-945. doi:10.1111/j.1439-0310.2008.01557.x

Dawkins, M. S. (2012). *Why animals matter : animal consciousness, animal welfare and human well-being*. Oxford: Oxford University Press.

de Passille, A. M., Jensen, M. B., Chapinal, N., & Rushen, J. (2010). Technical note: Use of accelerometers to describe gait patterns in dairy calves. *Journal of Dairy Science, 93*(7), 3287-3293. doi:10.3168/jds.2009-2758

de Waal, F. B. M. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences, 1224*(1), 191-206. doi:10.1111/j.1749-6632.2010.05912.x

Desborough, J. P. (2000). The stress response to trauma and surgery. *British Journal of Anaesthesia, 85*(1), 109-117. doi:DOI 10.1093/bja/85.1.109

Dettmer, A. M., Novak, M. A., Meyer, J. S., & Suomi, S. J. (2014). Population density-dependent hair cortisol concentrations in rhesus monkeys (Macaca mulatta). *Psychoneuroendocrinology, 42*, 59-67. doi:10.1016/j.psyneuen.2014.01.002

Diana, A., Salas, M., Pereboom, Z., Mendl, M., & Norton, T. (2021). A Systematic Review of the Use of Technology to Monitor Welfare in Zoo Animals: Is There Space for Improvement? *Animals, 11*(11), 3048. doi:10.3390/ani11113048

DiVincenti, L., & Wyatt, J. D. (2011). Pair Housing of Macaques in Research Facilities: A Science-Based Review of Benefits and Risks. *Journal of the American Association for Laboratory Animal Science, 50*(6), 856-863. Retrieved from <Go to ISI>://WOS:000306772000001

Dogra, P., & Vijayashankar, N. P. (2022). Dexamethasone Suppression Test. In *StatPearls*. Treasure Island (FL).

Drucker, C. B., Carlson, M. L., Toda, K., DeWind, N. K., & Platt, M. L. (2015). Non-invasive primate head restraint using thermoplastic masks. *Journal of Neuroscience Methods, 253*, 90-100. doi:10.1016/j.jneumeth.2015.06.013

Dulude-de Broin, F., Cote, S. D., Whiteside, D. P., & Mastromonaco, G. F. (2019). Faecal metabolites and hair cortisol as biological markers of HPA-axis activity in the Rocky mountain goat. *General and Comparative Endocrinology, 280*, 147-157. doi:10.1016/j.ygcen.2019.04.022

Duncan, I. J. H. (2006). The changing concept of animal sentience. *Applied Animal Behaviour Science, 100*(1-2), 11-19. doi:10.1016/j.applanim.2006.04.011

Duncan, I. J. H. (2019). Animal Welfare: A Brief History. In S. Hild & L. Schweitzer (Eds.), *Animal Welfare: From Science to Law* (pp. 13-19).

Fehlmann, G., O'Riain, M. J., Hopkins, P. W., O'Sullivan, J., Holton, M. D., Shepard, E. L. C., & King, A. J. (2017). Identification of behaviours from accelerometer data in wild social primates. *Animal Biotelemetry 5*(6), 1 - 11.

Fichtel, C., Kraus, C., Ganswindt, A., & Heistermann, M. (2007). Influence of reproductive season and rank on fecal glucocorticoid levels in free-ranging male Verreaux's sifakas (Propithecus verreauxi). *Hormones and Behavior, 51*(5), 640-648. doi:10.1016/j.yhbeh.2007.03.005

Field, A. E., Jones, C. L., Kelly, R., Marko, S. T., Kern, S. J., & Ricol, P. J. (2015). Measurement of Fecal Corticosterone Metabolites as a Predictor of the Habituation of Rhesus Macaques (Macaca mulatta) to Jacketing. *Journal of the American Association for Laboratory Animal Science, 54*(1), 59-65.

Flecknell, P. A., & Roughan, J. V. (2004). Assessing pain in animals - putting research into practice. *Animal Welfare, 13*, S71-S75. doi:10.1017/S0962728600014391

Fooden, J. (2000). Systematic Review of the Rhesus Macaques, *Macaca Mulatta* (Zimmermann, 1780). *Fieldiana Zoology, 96*.

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox, K. R. (1999). The influence of physical activity on mental well-being. *Public Health Nutrition, 2*(3a), 411-418. doi:10.1017/S1368980099000567

Fraser, D. (2003). Assessing animal welfare at the farm and group level: The interplay of science and values. *Animal Welfare, 12*(4), 433-443. Retrieved from <Go to ISI>://WOS:000186387100002

Fraser, D. (2008). Understanding Animal Welfare. *Acta Veterinaria Scandinavica, 50*(S1), 1 to 7. doi:10.1186/1751-0147-50-s1-s1

Fraser, D. (2009). Animal behaviour, animal welfare and the scientific study of affect. *Applied Animal Behaviour Science, 118*(3-4), 108-117. doi:10.1016/j.applanim.2009.02.020

Fries, E., Hesse, J., Hellhammer, J., & Hellhammer, D. H. (2005). A new view on hypocortisolism. *Psychoneuroendocrinology, 30*(10), 1010-1016. doi:10.1016/j.psyneuen.2005.04.006

Fujita, S., Noro, H., Nishikawa, K., Yumiba, T., Taniguchi, E., Ohashi, S., . . . Kimura, T. (2003). A New Parameter for Assessing Postoperative Recovery of Physical Activity Using an Accelerometer. *Surgery Today, 33*(9), 645-650. doi:10.1007/s00595-003-2582-7

Fuller, G., Margulis, S. W., & Santymire, R. (2011). The effectiveness of indigestible markers for identifying individual animal feces and their prevalence of use in North American zoos. *Zoo Biology, 30*(4), 379-398. doi:10.1002/zoo.20339

Ganswindt, A., Palme, R., Heistermann, M., Borragan, S., & Hodges, J. K. (2003). Non-invasive assessment of adrenocortical function in the male African elephant (Loxodonta africana) and its relation to musth. *Gen Comp Endocrinol, 134*(2), 156-166. doi:10.1016/s0016-6480(03)00251-x

Gao, W., Stalder, T., Foley, P., Rauh, M., Deng, H., & Kirschbaum, C. (2013). Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay. *J Chromatogr B Analyt Technol Biomed Life Sci, 928*, 1-8. doi:10.1016/j.jchromb.2013.03.008

Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., . . . Zwieg, A. S. (2007). Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. *Science, 316*(5822), 222-234. doi:doi:10.1126/science.1139247

Girard-Buttoz, C., Heistermann, M., Krummel, S., & Engelhardt, A. (2009). Seasonal and social influences on fecal androgen and glucocorticoid excretion in wild male long-tailed macaques (Macaca fascicularis). *Physiology & Behavior, 98*(1-2), 168-175. doi:10.1016/j.physbeh.2009.05.005

Gonzalez-de-la-Vara Mdel, R., Valdez, R. A., Lemus-Ramirez, V., Vazquez-Chagoyan, J. C., Villa-Godoy, A., & Romano, M. C. (2011). Effects of adrenocorticotropic hormone challenge and age on hair cortisol concentrations in dairy cattle. *Can J Vet Res, 75*(3), 216-221.

Gottlieb, D. H., Capitanio, J. P., & McCowan, B. (2013). Risk factors for stereotypic behavior and self-biting in rhesus macaques (Macaca mulatta): Animal's history, current environment, and personality. *American Journal of Primatology, 75*(10), 995-1008. doi:10.1002/ajp.22161

Gottlieb, D. H., Coleman, K., & McCowan, B. (2013). The effects of predictability in daily husbandry routines on captive rhesus macaques (Macaca mulatta). *Applied Animal Behaviour Science, 143*(2-4), 117-127. doi:10.1016/j.applanim.2012.10.010

Goymann, W. (2012). On the use of non-invasive hormone research in uncontrolled, natural environments: the problem with sex, diet, metabolic rate and the individual. *Methods in Ecology and Evolution, 3*(4), 757-765. doi:10.1111/j.2041-210X.2012.00203.x

Grandi, L. C., & Ishida, H. (2015). The Physiological Effect of Human Grooming on the Heart Rate and the Heart Rate Variability of Laboratory Non-Human Primates: A Pilot Study in Male Rhesus Monkeys. *Front Vet Sci, 2*, 50. doi:10.3389/fvets.2015.00050

Grandi, L. C., Roda, F., & Ishida, H. (2015). Physiological Effect of Sweeping Grooming Movements in Macaque Monkey: Preliminary Data. *Journal of Primatology, 4*(2).

Greco, B. J., Meehan, C. L., Heinsius, J. L., & Mench, J. A. (2017). Why pace? The influence of social, housing, management, life history, and demographic characteristics on locomotor stereotypy in zoo elephants. *Applied Animal Behaviour Science, 194*, 104-111. doi:10.1016/j.applanim.2017.05.003

Green, T., & Mellor, D. (2011). Extending ideas about animal welfare assessment to include 'quality of life' and related concepts. *New Zealand Veterinary Journal, 59*(6), 263-271. doi:10.1080/00480169.2011.610283

Hadinger, U., Haymerle, A., Knauer, F., Schwarzenberger, F., & Walzer, C. (2015). Faecal cortisol metabolites to assess stress in wildlife: evaluation of a field method in free-ranging chamois. *Methods in Ecology and Evolution, 6*(11), 1349-1357. doi:10.1111/2041-210x.12422

Hamalainen, A., Heistermann, M., Fenosoa, Z. S., & Kraus, C. (2014). Evaluating capture stress in wild gray mouse lemurs via repeated fecal sampling: method validation and the influence of prior experience and handling protocols on stress responses. *Gen Comp Endocrinol, 195*, 68-79. doi:10.1016/j.ygcen.2013.10.017

Hannibal, D. L., Bliss-Moreau, E., Vandeleest, J., McCowan, B., & Capitanio, J. (2017). Laboratory rhesus macaque social housing and social changes: Implications for research. *Am J Primatol, 79*(1), 1-14. doi:10.1002/ajp.22528

Harrison, R. (1964). *Animal Machines*. London: Vincent Stuart.
Hawkins, P., Morton, D. B., Bevan, R., Heath, K., Kirkwood, J., Pearce, P., . . . Webb, A. (2004). Husbandry refinements for rats, mice, dogs and non-human primates used in telemetry procedures - Seventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement, Part B. *Laboratory Animals, 38*(1), 1-10.
Heimburge, S., Kanitz, E., & Otten, W. (2019). The use of hair cortisol for the assessment of stress in animals. *General and Comparative Endocrinology, 270*, 10-17. doi:10.1016/j.ygcen.2018.09.016
Heimburge, S., Kanitz, E., Tuchscherer, A., & Otten, W. (2020). Is it getting in the hair? - Cortisol concentrations in native, regrown and segmented hairs of cattle and pigs after repeated ACTH administrations. *General and Comparative Endocrinology, 295*. doi:10.1016/j.ygcen.2020.113534
Heistermann, M., Ademmer, C., & Kaumanns, W. (2004). Ovarian cycle and effect of social changes on adrenal and ovarian function in Pygathrix nemaeus. *International Journal of Primatology, 25*(3), 689-708. doi:DOI 10.1023/B:IJOP.0000023581.17889.0f
Heistermann, M., Finke, M., & Hodges, J. K. (1995). Assessment of Female Reproductive Status in Captive-Housed Hanuman Langurs (Presbytis-Entellus) by Measurement of Urinary and Fecal Steroid-Excretion Patterns. *American Journal of Primatology, 37*(4), 275-284. doi:DOI 10.1002/ajp.1350370402
Heistermann, M., Palme, R., & Ganswindt, A. (2006). Comparison of different enzymeimmunoassays for assessment of adrenocortical activity in primates based on fecal analysis. *American Journal of Primatology, 68*(3), 257-273. doi:10.1002/ajp.20222
Hemsworth, P. H., Mellor, D. J., Cronin, G. M., & Tilbrook, A. J. (2015). Scientific assessment of animal welfare. *New Zealand Veterinary Journal, 63*(1), 24-30. doi:10.1080/00480169.2014.966167
Hendrickx, H. H. L., Safar, P., & Miller, A. (1984). Delayed recovery of behavior after anesthesia in rats. *Resuscitation, 12*(3), 213-221. doi:10.1016/0300-9572(84)90008-x
Hennessy, M. B., Davis, H. N., Williams, M. T., Mellott, C., & Douglas, C. W. (1997). Plasma cortisol levels of dogs at a county animal shelter. *Physiology & Behavior, 62*(3), 485-490. doi:10.1016/s0031-9384(97)80328-9
Higham, J. P., Heistermann, M., & Maestripieri, D. (2013). The endocrinology of male rhesus macaque social and reproductive status: a test of the challenge and social stress hypotheses. *Behav Ecol Sociobiol, 67*(1), 19-30. doi:10.1007/s00265-012-1420-6
Higham, J. P., Vitale, A. B., Rivera, A. M., Ayala, J. E., & Maestripieri, D. (2010). Measuring salivary analytes from free-ranging monkeys. *Physiol Behav, 101*(5), 601-607. doi:10.1016/j.physbeh.2010.09.003
Hildebrandt, I. J., Su, H., & Weber, W. A. (2008). Anesthesia and other considerations for in vivo imaging of small animals. *Ilar Journal, 49*(1), 17-26. doi:10.1093/ilar.49.1.17
Hoffman, C. L., Higham, J. P., Heistermann, M., Coe, C. L., Prendergast, B. J., & Maestripieri, D. (2011). Immune function and HPA axis activity in free-ranging rhesus macaques. *Physiol Behav, 104*(3), 507-514. doi:10.1016/j.physbeh.2011.05.021
Hohlbaum, K., Bert, B., Dietze, S., Palme, R., Fink, H., & Thöne-Reineke, C. (2017). Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice—Assessing the degree of distress. *Plos One, 12*(6), e0179588. doi:10.1371/journal.pone.0179588
Hohlbaum, K., Bert, B., Dietze, S., Palme, R., Fink, H., & Thöne-Reineke, C. (2018). Impact of repeated anesthesia with ketamine and xylazine on the well-being of C57BL/6JRj mice. *Plos One, 13*(9), e0203559. doi:10.1371/journal.pone.0203559
Home Office. (2014a). *Code of Practice for the Housing and Care of Animals Bred, Supplied or Used for Scientifc Purposes*. London: Stationary Office.
Home Office. (2014b). *Guidance of the Operation of the Animals (Scientific Procedures) Act 1986*. Retrieved from

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662364/Guidance_on_the_Operation_of_ASPA.pdf
Home Office. (2015). *The Harm-Benefit Analysis Process: New Project Licence Applications*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/487914/Harm_Benefit_Analysis__2_.pdf
Home Office. (2022). *Annual Statistics of Scientific Procedures on Living Animals, Great Britain 2021*. www.gov.uk/official-documents
Honess, P., & Wolfensohn, S. (2010). The Extended Welfare Assessment Grid: A Matrix for the Assessment of Welfare and Cumulative Suffering in Experimental Animals. *Alternatives to Laboratory Animals, 38*(3), 205-212. doi:10.1177/026119291003800304
Howell, C. P., & Cheyne, S. M. (2019). Complexities of Using Wild versus Captive Activity Budget Comparisons for Assessing Captive Primate Welfare. *Journal of Applied Animal Welfare Science, 22*(1), 78-96. doi:10.1080/10888705.2018.1500286
Hubrecht, R. C. (2014). *The Welfare of Animals Used in Research: Practice and ethics*. Chichester: Wily Blackwell.
Hunt, K. E., & Wasser, S. K. (2003). Effect of long-term preservation methods on fecal glucocorticoid concentrations of grizzly bear and African elephant. *Physiological and Biochemical Zoology, 76*(6), 918-928. doi:Doi 10.1086/380209
Hurst, M. W., Jenkins, C. D., & Rose, R. M. (1976). The relation of psychological stress to onset of medical illness. *Annu Rev Med, 27*, 301-312. doi:10.1146/annurev.me.27.020176.001505
InformedHealth.org. (2019). What is an electrocardiogram (ECG)? Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK536878/?report=reader
Ipsos MORI. (2016). *Public attitudes to animal research in 2016*. Retrieved from https://www.ipsos.com/sites/default/files/publication/1970-01/sri-public-attitudes-to-animal-research-2016.pdf
Ipsos MORI. (2018). *Public attitudes to animal research in 2018*. Retrieved from https://www.ipsos.com/sites/default/files/ct/news/documents/2019-05/18-040753-01_ols_public_attitudes_to_animal_research_report_v3_191118_public.pdf
Jackson, M. N., Truelove, M. A., Williams, K., Chen, J. D., Moore, R. H., Wood, J. S., . . . Mollie, B. (2023). Effects of pair housing on behavior, cortisol, and clinical outcomes during quarantine-like procedures for rhesus macaques (Macaca mulatta). *Journal of Medical Primatology*. doi:10.1111/jmp.12635
Jacobsen, K. R., Kalliokoski, O., Teilmann, A. C., Hau, J., & Abelson, K. S. P. (2012a). The effect of isoflurane anaesthesia and vasectomy on circulating corticosterone and ACTH in BALB/c mice. *General and Comparative Endocrinology, 179*(3), 406-413. doi:10.1016/j.ygcen.2012.09.012
Jacobsen, K. R., Kalliokoski, O., Teilmann, A. C., Hau, J., & Abelson, K. S. P. (2012b). Postsurgical Food and Water Consumption, Fecal Corticosterone Metabolites, and Behavior Assessment as Noninvasive Measures of Pain in Vasectomized BALB/c Mice. *Journal of the American Association for Laboratory Animal Science, 51*(1), 69-75.
Jarczok, M. N., Koenig, J., Shively, C. A., & Thayer, J. F. (2018). Behavioral depression is associated with increased vagally mediated heart rate variability in adult female cynomolgus monkeys (Macaca fascicularis). *International Journal of Psychophysiology, 131*, 139-143. doi:10.1016/j.ijpsycho.2017.11.004
Jennings, M., Prescott, M. J., Buchanan-Smith, H. M., Gamble, M. R., Gore, M., Hawkins, P., . . . Buist, D. (2009). Refinements in husbandry, care and common procedures for non-human primates:Ninth report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Laboratory Animals, 43*(1_suppl), 1-47. doi:10.1258/la.2008.007143

Johnston, J. M., Cohen, Y. E., Shirley, H., Tsunada, J., Bennur, S., Christison-Lagay, K., & Veeder, C. L. (2016). Recent refinements to cranial implants for rhesus macaques (Macaca mulatta). *Lab Animal, 45*(5), 180-186. doi:10.1038/laban.997

Jonckheer-Sheehy, V. S. M., Vinke, C. M., & Ortolani, A. (2012). Validation of a Polar (R) human heart rate monitor for measuring heart rate and heart rate variability in adult dogs under stationary conditions. *Journal of Veterinary Behavior-Clinical Applications and Research, 7*(4), 205-212. doi:10.1016/j.jveb.2011.10.006

Jones, S., Dowling-Guyer, S., Patronek, G. J., Marder, A. R., Segurson D'Arpino, S., & McCobb, E. (2014). Use of Accelerometers to Measure Stress Levels in Shelter Dogs. *Journal of Applied Animal Welfare Science, 17*(1), 18-28. doi:10.1080/10888705.2014.856241

Jonker, L. T., Hendriks, S., Lahr, M. M., Van Munster, B. C., De Bock, G. H., & Van Leeuwen, B. L. (2020). Postoperative recovery of accelerometer-based physical activity in older cancer patients. *European Journal of Surgical Oncology, 46*(11), 2083-2090. doi:10.1016/j.ejso.2020.06.012

Kamps, A. W., Molenmaker, M., Kemperman, R., van der Veen, B. S., Bocca, G., & Veeger, N. J. (2014). Children with asthma have significantly lower long-term cortisol levels in their scalp hair than healthy children. *Acta Paediatr, 103*(9), 957-961. doi:10.1111/apa.12685

Karim, N., Hasan, J. A., & Ali, S. S. (2011). Heart rate variability - a review. *Journal of Basic and Applied Sciences 7*(1), 71-77.

Kelz, M. B., & Mashour, G. A. (2019). The Biology of General Anesthesia from Paramecium to Primate. *Curr Biol, 29*(22), R1199-R1210. doi:10.1016/j.cub.2019.09.071

Kher, R. (2019). Signal Processing Techniques for Removing Noise from ECG Signals. *Journal of Biomedical Engineering and Research, 3*(101), 1-9.

Kikusui, T., Winslow, J. T., & Mori, Y. (2006). Social buffering: relief from stress and anxiety. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361*(1476), 2215-2228. doi:10.1098/rstb.2006.1941

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation, 15*(3), 235-245. doi:10.30773/pi.2017.08.17

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary Cortisol in Psychoneuroendocrine Research - Recent Developments and Applications. *Psychoneuroendocrinology, 19*(4), 313-333. doi:Doi 10.1016/0306-4530(94)90013-2

Kirschbaum, C., Tietze, A., Skoluda, N., & Dettenborn, L. (2009). Hair as a retrospective calendar of cortisol production—Increased cortisol incorporation into hair in the third trimester of pregnancy. *Psychoneuroendocrinology, 34*(1), 32-37. doi:10.1016/j.psyneuen.2008.08.024

Koolhaas, J. M., de Boer, S. F., Coppens, C. M., & Buwalda, B. (2010). Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology, 31*(3), 307-321. doi:10.1016/j.yfrne.2010.04.001

Koren, L., Bryan, H., Matas, D., Tinman, S., Fahlman, A., Whiteside, D., . . . Wynne-Edwards, K. (2019). Towards the validation of endogenous steroid testing in wildlife hair. *Journal of Applied Ecology, 56*(3), 547-561. doi:10.1111/1365-2664.13306

Kovacs, L., Jurkovich, V., Bakony, M., Szenci, O., Poti, P., & Tozser, J. (2014). Welfare implication of measuring heart rate and heart rate variability in dairy cattle: literature review and conclusions for future research. *Animal, 8*(2), 316-330. doi:10.1017/S1751731113002140

Kovacs, L., Kezer, F. L., Jurkovich, V., Kulcsar-Huszenicza, M., & Tozser, J. (2015). Heart Rate Variability as an Indicator of Chronic Stress Caused by Lameness in Dairy Cows. *Plos One, 10*(8). doi:10.1371/journal.pone.0134792

Kremer, L., Holkenborg, S. E. J. K., Reimert, I., Bolhuis, J. E., & Webb, L. E. (2020). The nuts and bolts of animal emotion. *Neuroscience and Biobehavioral Reviews, 113*, 273-286. doi:10.1016/j.neubiorev.2020.01.028

Kuhn, M. (2022). caret: Classification and Regression Training. Retrieved from https://CRAN.R-project.org/package=caret

Ladewig, J. (2000). Chronic Intermittent Stress: A model for the study of long-term stressors. In G. P. Moberg & J. A. Mench (Eds.), *The Biology of Animal Stress: Basic principles and implications for animal welfare* (pp. 159-169). Wallingford, Oxon, UK: CABI Publishing.

Lee, T., Jarome, T., Li, S. J., Kim, J. J., & Helmstetter, F. J. (2009). Chronic stress selectively reduces hippocampal volume in rats: a longitudinal magnetic resonance imaging study. *Neuroreport, 20*(17), 1554-1558. doi:10.1097/WNR.0b013e328332bb09

Lemon, R. N. (2018). Applying the 3Rs to neuroscience research involving nonhuman primates. *Drug Discovery Today, 23*(9), 1574-1577. doi:10.1016/j.drudis.2018.05.002

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y., Murgatroyd, M., & Papastamatiou, Y. P. (2017). Analysis of animal accelerometer data using hidden Markov models. *Methods in Ecology and Evolution, 8*(2), 161-173. doi:10.1111/2041-210x.12657

Lilly, A. A., Mehlman, P. T., & Higley, J. D. (1999). Trait-like immunological and hematological measures in female rhesus across varied environmental conditions. *American Journal of Primatology, 48*(3), 197-223. doi:Doi 10.1002/(Sici)1098-2345(1999)48:3

Long, J. A. (2020). _jtools: Analysis and Presentation of Social Scientific Data_. R package version 2.1.0. Retrieved from https://cran.r-project.org/package=jtools

Loomes, K., & Louro, L. F. (2021). Recovery of horses from general anaesthesia: A systematic review (2000-2020) of risk factors and influence of interventions during the recovery period. *Equine Veterinary Journal*. doi:10.1111/evj.13517

Lutz, C., Well, A., & Novak, M. (2003). Stereotypic and self-injurious behavior in rhesus macaques: A survey and retrospective analysis of environment and early experience. *American Journal of Primatology, 60*(1), 1-15. doi:10.1002/ajp.10075

Lutz, C. K., Menard, M. T., Rosenberg, K., Meyer, J. S., & Novak, M. A. (2019). Alopecia in rhesus macaques (Macaca mulatta): Association with pregnancy and chronic stress. *Journal of Medical Primatology, 48*(4), 251-256. doi:10.1111/jmp.12419

Lutz, C. K., Tiefenbacher, S., Jorgensen, M. J., Meyer, J. S., & Novak, M. A. (2000). Techniques for collecting saliva from awake, unrestrained, adult monkeys for cortisol assay. *American Journal of Primatology, 52*(2), 93-99. doi:Doi 10.1002/1098-2345(200010)52:2<93::Aid-Ajp3>3.3.Co;2-2

Macbeth, B. J., Cattet, M. R. L., Stenhouse, G. B., Gibeau, M. L., & Janz, D. M. (2010). Hair cortisol concentration as a noninvasive measure of long-term stress in free-ranging grizzly bears (Ursus arctos): considerations with implications for other wildlife. *Canadian Journal of Zoology-Revue Canadienne De Zoologie, 88*(10), 935-949. doi:10.1139/Z10-057

Magden, E. R., Mansfield, K. G., Simmons, J. H., & Abee, C. R. (2015). Nonhuman Primates *In: Fox. J, Anderson. L, Otto. G, Pritchett-Corning. K, Whary. M (editors). Laboratory Animal Medicine 3rd edition Academic Press, 2015*, Chapter 17, p. 771 - 930.

Mann, T. M., Williams, K. E., Pearce, P. C., & Scott, E. A. (2005). A novel method for activity monitoring in small non-human primates. *Lab Anim, 39*(2), 169-177. doi:10.1258/0023677053739783

Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., . . . Rushworth, M. F. S. (2011). Diffusion-Weighted Imaging Tractography-Based Parcellation of the Human Parietal Cortex and Comparison with Human and Macaque Resting-State Functional Connectivity. *Journal of Neuroscience, 31*(11), 4087-4100. doi:10.1523/Jneurosci.5102-10.2011

Martin, P., & Bateson, P. P. G. (2007). *Measuring behaviour : an introductory guide* (3rd ed.). Cambridge ; New York: Cambridge University Press.

Marty, P. R., Hodges, K., Heistermann, M., Agil, M., & Engelhardt, A. (2017). Is social dispersal stressful? A study in male crested macaques (Macaca nigra). *Hormones and Behavior, 87*, 62-68. doi:10.1016/j.yhbeh.2016.10.018

Mason, G., & Mendl, M. (1993). Why is there no simple way of measuring animal welfare? *Animal Welfare, 2*(4), 301-319.

Mason, G. J., Cooper, J., & Clarebrough, C. (2001). Frustrations of fur-farmed mink. *Nature, 410*(6824), 35-36. doi:10.1038/35065157

Mason, G. J., & Latham, N. R. (2004). Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? *Animal Welfare, 13*, S57-S69.

Mason, S., Premereur, E., Pelekanos, V., Emberton, A., Honess, P., & Mitchell, A. S. (2019). Effective chair training methods for neuroscience research involving rhesus macaques (Macaca mulatta). *Journal of Neuroscience Methods, 317*, 82-93. doi:10.1016/j.jneumeth.2019.02.001

Mather, M., & Thayer, J. F. (2018). How heart rate variability affects emotion regulation brain networks. *Current Opinion in Behavioral Sciences, 19*, 98-104. doi:10.1016/j.cobeha.2017.12.017

Matteri, R. L., Carroll, J.A and Dyer, C.J. (2000). Neuroendocrine Responses to Stress. In G. P. Moberg & J. A. Mench (Eds.), *The Biology of Animal Stress: Basic principles and implications for animal welfare* (pp. 43-76). Wallingford, Oxon, UK: CABI Publishing.

McFarland, R., Hetem, R. S., Fuller, A., Mitchell, D., Henzi, S. P., & Barrett, L. (2013). Assessing the reliability of biologger techniques to measure activity in a free-ranging primate. *Animal Behaviour, 85*(4), 861-866. doi:10.1016/j.anbehav.2013.02.005

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276-282. doi:DOI 10.11613/bm.2012.031

McMillan, J. L., Bloomsmith, M. A., & Prescott, M. J. (2017). An International Survey of Approaches to Chair Restraint of Nonhuman Primates. *Comp Med, 67*(5), 442-451.

Meek, S., & Morris, F. (2002). ABC of clinical electrocardiography - Introduction. I - Leads, rate, rhythm, and cardiac axis. *British Medical Journal, 324*(7334), 415-418. doi:DOI 10.1136/bmj.324.7334.415

Mellor, D. (2015). Enhancing animal welfare by creating opportunities for positive affective engagement. *New Zealand Veterinary Journal, 63*(1), 3-8. doi:10.1080/00480169.2014.926799

Mellor, D. (2016). Updating Animal Welfare Thinking: Moving beyond the “Five Freedoms” towards “A Life Worth Living”. *Animals, 6*(3), 21. doi:10.3390/ani6030021

Mellor, D. J. (2017). Operational Details of the Five Domains Model and Its Key Applications to the Assessment and Management of Animal Welfare. *Animals, 7*(12), 60. doi:10.3390/ani7080060

Mellor, D. J., Beausoleil, N. J., Littlewood, K. E., McLean, A. N., McGreevy, P. D., Jones, B., & Wilkins, C. (2020). The 2020 Five Domains Model: Including Human-Animal Interactions in Assessments of Animal Welfare. *Animals (Basel), 10*(10). doi:10.3390/ani10101870

Mellor, D. J., Patterson-Kane, E., Stafford, K. J., & Universities Federation for Animal Welfare. (2009). *The sciences of animal welfare*. Oxford ; Ames, Iowa: Wiley-Blackwell.

Mellor, D. J., & Reid, C. S. W. (1994). *Concepts of animal well-being and predicting the impact of procedures on experimental animals.* Paper presented at the Improving the Well-being of Animals in the Research Environment, Glen Osmond, Australia.

Menargues, A., Urios, V., & Mauri, M. (2008). Welfare assessment of captive Asian elephants (Elephas maximus) and Indian rhinoceros (Rhinoceros unicornis) using salivary cortisol measurement. *Animal Welfare, 17*(3), 305-312. doi:10.1017/S096272860003222X

Mendez, J. C., Perry, B. A. L., Heppenstall, R. J., Mason, S., & Mitchell, A. S. (2022). Openness about animal research increases public support. *Nature Neuroscience, 25*(4), 401-403. doi:10.1038/s41593-022-01039-z

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences, 277*(1696), 2895-2904. doi:10.1098/rspb.2010.0303

Mendl, M., & Paul, E. S. (2004). Consciousness, emotion and animal welfare: insights from cognitive science. *Animal Welfare, 13*, S17-S25. doi:10.1017/S0962728600014330

Meyer, J., Novak, M., Hamel, A., & Rosenberg, K. (2014). Extraction and Analysis of Cortisol from Human and Monkey Hair. *Jove-Journal of Visualized Experiments*(83). doi:10.3791/50882

Meyer, J. S., & Novak, M. A. (2012). Minireview: Hair cortisol: a novel biomarker of hypothalamic-pituitary-adrenocortical activity. *Endocrinology, 153*(9), 4120-4127. doi:10.1210/en.2012-1226

Miller, L. J., Vicino, G. A., Sheftel, J., & Lauderdale, L. K. (2020). Behavioral Diversity as a Potential Indicator of Positive Animal Welfare. *Animals, 10*(7), 1211. doi:10.3390/ani10071211

Mitchell, A. S., Hartig, R., Basso, M. A., Jarrett, W., Kastner, S., & Poirier, C. (2021). International primate neuroscience research regulation, public engagement and transparency opportunities. *Neuroimage, 229*, 117700. doi:10.1016/j.neuroimage.2020.117700

Mitchell, A. S., Thiele, A., Petkov, C. I., Roberts, A., Robbins, T. W., Schultz, W., & Lemon, R. (2018). Continued need for non-human primate neuroscience research. *Current Biology, 28*(20), R1186-R1187. doi:DOI 10.1016/j.cub.2018.09.029

Moberg, G. P. (2000). Biological Responses to Stress: Implications for Animal Welfare. In G. P. Moberg & J. A. Mench (Eds.), *The Biology of Animal Stress: Basic principles and implications for animal welfare* (pp. 1-21). Wallingford, Oxon, UK: CABI Publishing.

Molony, V., Kent, J. E., & Robertson, I. S. (1993). Behavioural responses of lambs of three ages in the first three hours after three methods of castration and tail docking. *Research in Veterinary Science, 55*(2), 236-245. doi:10.1016/0034-5288(93)90087-v

Mormede, P., Andanson, S., Auperin, B., Beerda, B., Guemene, D., Malmkvist, J., . . . Veissier, I. (2007). Exploration of the hypothalamic-pituitary-adrenal function as a tool to evaluate animal welfare. *Physiology & Behavior, 92*(3), 317-339. doi:10.1016/j.physbeh.2006.12.003

Morrow-Tesch, J. L., McGlone, J. J., & Norman, R. L. (1993). Consequences of Restraint Stress on Natural-Killer-Cell Activity, Behavior, and Hormone Levels in Rhesus Macaques (Macaca-Mulatta). *Psychoneuroendocrinology, 18*(5-6), 383-395. doi:10.1016/0306-4530(93)90013-B

Morton, D. B., Hawkins, P., Bevan, R., Heath, K., Kirkwood, J., Pearce, P., . . . Refinement, J. W. G. (2003). Refinements in telemetry procedures - Seventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement, part A. *Laboratory Animals, 37*(4), 261-299. doi:Doi 10.1258/002367703322389861

Mostl, E., & Palme, R. (2002). Hormones as indicators of stress. *Domestic Animal Endocrinology, 23*(1-2), 67-74. doi:10.1016/S0739-7240(02)00146-7

Mueller, C., Jenni-Eiermann, S., Blondel, J., Perret, P., Caro, S. P., Lambrechts, M., & Jenni, L. (2006). Effect of human presence and handling on circulating corticosterone levels in breeding blue tits (Parus caeruleus). *General and Comparative Endocrinology, 148*(2), 163-171. doi:10.1016/j.ygcen.2006.02.012

Mukherjee, J., Mohapatra, S. S., Jana, S., Das, P. K., Ghosh, P. R., Das, K., & Banerjee, D. (2020). A study on the electrocardiography in dogs: Reference values and their comparison among breeds, sex, and age groups. *Veterinary World, 13*(10), 2216-2220. doi:10.14202/vetworld.2020.2216-2220

National Human Genome Research Institute. (2010). Why Mouse Matters. Retrieved from https://www.genome.gov/10001345/importance-of-mouse-genome

National Research Council. (2008). *Recognition and alleviation of distress in laboratory animals*. Washington, D.C.: National Academies Press.

NC3Rs. (2015). The Macaque Website. Retrieved from https://macaques.nc3rs.org.uk/about-macaques/rhesus-macaque
NC3Rs. (2021). Guidance: Anaesthesia. Retrieved from https://www.nc3rs.org.uk/3rs-resources/anaesthesia
Neethirajan, S., Reimert, I., & Kemp, B. (2021). Measuring Farm Animal Emotions—Sensor-Based Approaches. *Sensors, 21*(2), 553. doi:10.3390/s21020553
Novak, M. A., Hamel, A. F., Coleman, K., Lutz, C. K., Worlein, J., Menard, M., . . . Meyer, J. S. (2014). Hair loss and hypothalamic-pituitary-adrenocortical axis activity in captive rhesus macaques (Macaca mulatta). *J Am Assoc Lab Anim Sci, 53*(3), 261-266.
Novak, M. A., Hamel, A. F., Kelly, B. J., Dettmer, A. M., & Meyer, J. S. (2013). Stress, the HPA axis, and nonhuman primate well-being: A review. *Appl Anim Behav Sci, 143*(2-4), 135-149. doi:10.1016/j.applanim.2012.10.012
Novak, M. A., & Suomi, S. J. (1989). Psychological Well-Being of Captive Primates. *Science and Animals : Addressing Contemporary Issues*, 5-12. Retrieved from <Go to ISI>://WOS:A1989BS37T00001
Ostner, J., Heistermann, M., & Schulke, O. (2008). Dominance, aggression and physiological stress in wild male Assamese macaques (Macaca assamensis). *Hormones and Behavior, 54*(5), 613-619. doi:10.1016/j.yhbeh.2008.05.020
Otovic, P., & Hutchinson, E. (2015). Limits to Using HPA Axis Activity as an Indication of Animal Welfare. *Altex-Alternatives to Animal Experimentation, 32*(1), 41-50. doi:10.14573/altex.1406161.
Palme, R. (2012). Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animals. *Animal Welfare, 21*(3), 331-337. doi:10.7120/09627286.21.3.331
Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiology & Behavior, 199*, 229-243. doi:10.1016/j.physbeh.2018.11.021
Papailiou, A., Sullivan, E., & Cameron, J. L. (2008). Behaviors in rhesus monkeys (Macaca mulatta) associated with activity counts measured by accelerometer. *Am J Primatol, 70*(2), 185-190. doi:10.1002/ajp.20476
Papich, M. G. (2021). Prednisone. In *Papich Handbook of Veterinary Drugs* (5 ed., pp. 771 - 773): Saunders Elsevier.
Patterson, S. M., Krantz, D. S., Montgomery, L. C., Deuster, P. A., Hedges, S. M., & Nebel, L. E. (1993). Automated physical activity monitoring: validation and comparison with physiological and self-report measures. *Psychophysiology, 30*(3), 296-305. doi:10.1111/j.1469-8986.1993.tb03356.x
Paul, E. S., Harding, E. J., & Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews, 29*(3), 469-491. doi:10.1016/j.neubiorev.2005.01.002
Pearson, B. L., Judge, P. G., & Reeder, D. M. (2008). Effectiveness of saliva collection and enzyme-immunoassay for the quantification of cortisol in socially housed baboons. *American Journal of Primatology, 70*(12), 1145-1151. doi:10.1002/ajp.20613
Percie du Sert, N., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., . . . Würbel, H. (2020). Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. *PLOS Biology, 18*(7), e3000411. doi:10.1371/journal.pbio.3000411
Peric, T., Comin, A., Corazzin, M., Montillo, M., Canavese, F., Stebel, M., & Prandi, A. (2018). Hair cortisol concentrations in New Zealand white rabbits subjected to surgery. *Animal Welfare, 27*(1), 13-20. doi:10.7120/09627286.27.1.013
Permar, T., & Williamson, K. (2021). History of the Electrocardiogram. In W. J. Brady, M. J. Lipinski, A. E. Darby, M. C. Bond, N. P. Charlton, K. Hudson, & K. Williamson (Eds.), *Electrocardiogram in Clinical Medicine* (pp. 13-18). Oxford: John Wiley & Sons Ltd.

Perretta, G. (2009). Non-Human Primate Models in Neuroscience Research. *Scandinavian Journal of Laboratory Animal Science, 36*(1), 77-85.

Perry, B. A. L., Mason, S., Nacef, J., Waddle, A., Hynes, B., Bergmann, C., . . . Mitchell, A. S. (2021). Protective cranial implant caps for macaques. *Journal of Neuroscience Methods, 348*, 108992. doi:10.1016/j.jneumeth.2020.108992

Pfefferle, D., Plümer, S., Burchardt, L., Treue, S., & Gail, A. (2018). Assessment of stress responses in rhesus macaques (Macaca mulatta) to daily routine procedures in system neuroscience based on salivary cortisol concentrations. *Plos One, 13*(1), e0190190. doi:10.1371/journal.pone.0190190

Phillips, K. A., Bales, K. L., Capitanio, J. P., Conley, A., Czoty, P. W., 't Hart, B. A., . . . Voytko, M. L. (2014). Why Primate Models Matter. *American Journal of Primatology, 76*(9), 801-827. doi:10.1002/ajp.22281

Piccione, G., Giannetto, C., Casella, S., Caola, G. . (2010). Daily locomotor activity in five domestic animals. *Animal Biology, 60*, 15-24.

Pickard, J. (2013). REVIEW OF THE ASSESSMENT OF CUMULATIVE SEVERITY AND LIFETIME EXPERIENCE IN NON-HUMAN PRIMATES USED IN NEUROSCIENCE RESEARCH. *Animal Procedures Committee (APC), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/261687/cs_nhp_review_FINAL_2013_corrected.pdf*.

Platt, J. P., Simon, B. T., Coleman, M., Martinez, E. A., Lepiz, M. A., & Watts, A. E. (2018). The effects of multiple anaesthetic episodes on equine recovery quality. *Equine Veterinary Journal, 50*(1), 111-116. doi:10.1111/evj.12728

Poirier, C., & Bateson, M. (2017). Pacing stereotypies in laboratory rhesus macaques: Implications for animal welfare and the validity of neuroscientific findings. *Neurosci Biobehav Rev*. doi:10.1016/j.neubiorev.2017.09.010

Poirier, C., Oliver, C. J., Bueno, J. C., Flecknell, P., & Bateson, M. (2019). Pacing behaviour in laboratory macaques is an unreliable indicator of acute stress. *Scientific Reports, 9*. doi:10.1038/s41598-019-43695-5

Pol, F., Courboulay, V., Cotte, J. P., Martrenchar, A., Hay, M., & Mormede, P. (2002). Urinary cortisol as an additional tool to assess the welfare of pregnant sows kept in two types of housing. *Veterinary Research, 33*(1), 13-22. doi:10.1051/vetres:2001002

Prescott, M. J. (2010). Ethics of primate use. *Advances in Science & Research* (5), 11-22. doi:10.5194/asr-5-11-2010

Prescott, M. J., Langermans, J. A., & Ragan, I. (2017). Applying the 3Rs to non-human primate research: Barriers and solutions. *Drug Discovery Today: Disease Models, 23*, 51-56. doi:10.1016/j.ddmod.2017.11.001

Prescott, M. J., & Lidster, K. (2017). Improving quality of science through better animal welfare: the NC3Rs strategy. *Lab Animal, 46*(4), 152-156. doi:https://doi.org/10.1038/laban.1217

Prete, A., Yan, Q., Al-Tarrah, K., Akturk, H. K., Prokop, L. J., Alahdab, F., . . . Bancos, I. (2018). The cortisol stress response induced by surgery: A systematic review and meta-analysis. *Clin Endocrinol (Oxf), 89*(5), 554-567. doi:10.1111/cen.13820

Proctor, H. (2012). Animal Sentience: Where Are We and Where Are We Heading? *Animals (Basel), 2*(4), 628-639. doi:10.3390/ani2040628

Proctor, H. S., Carder, G., & Cornish, A. R. (2013). Searching for Animal Sentience: A Systematic Review of the Scientific Literature. *Animals (Basel), 3*(3), 882-906. doi:10.3390/ani3030882

Pulopulos, M. M., Vanderhasselt, M. A., & De Raedt, R. (2018). Association between changes in heart rate variability during the anticipation of a stressful situation and the stress-induced cortisol response. *Psychoneuroendocrinology, 94*, 63-71. doi:10.1016/j.psyneuen.2018.05.004

Qin, D., Rizak, J., Chu, X., Li, Z., Yang, S., Lu, L., . . . Hu, X. (2015). A spontaneous depressive pattern in adult female rhesus macaques. *Sci Rep, 5*, 11267. doi:10.1038/srep11267
R Core Team. (2013). R: A language and environment for statistical computing.
R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/.
Rains, P. C., Rampersad, N., De Lima, J., Murrell, D., Kinchington, D., Lee, J. W., . . . Donaghue, K. C. (2009). Cortisol response to general anaesthesia for medical imaging in children. *Clinical Endocrinology, 71*(6), 834-839. doi:10.1111/j.1365-2265.2009.03591.x
Ratcliffe, A. M., Zhai, B., Guan, Y., Jackson, D. G., Sneyd, J. R., Minto, G., . . . Spinney, S. (2021). Patient-centred measurement of recovery from day-case surgery using wrist worn accelerometers: a pilot and feasibility study*. *Anaesthesia, 76*(6), 785-797. doi:10.1111/anae.15267
Rauch, E., Bergmann, S., Hagn, A., Meixensperger, J., Reese, S., Palme, R., & Erhard, M. H. (2014). Age-dependent baseline values of faecal cortisol metabolites in the American mink (Neovison vison) under semi- natural housing conditions. *Journal of Animal Physiology and Animal Nutrition, 98*(3), 497-503. doi:10.1111/jpn.12098
Reed, M. J., Robertson, C. E., & Addison, P. S. (2005). Heart rate variability measurements and the prediction of ventricular arrhythmias. *Qjm-an International Journal of Medicine, 98*(2), 87-95. doi:10.1093/qjmed/hci018
Reeder, D. M., & Kramer, K. M. (2005). Stress in free-ranging mammals: Integrating physiology, ecology, and natural history. *Journal of Mammalogy, 86*(2), 225-235. doi:10.1644/Bhe-003.1
Rennie, A. E., & Buchanan-Smith, H. M. (2006a). Refinement of the use of non-human primates in scientific research. Part I: the influence of humans. *Animal Welfare, 15*(3), 203-213. Retrieved from <Go to ISI>://WOS:000239709100001
Rennie, A. E., & Buchanan-Smith, H. M. (2006b). Refinement of the use of non-human primates in scientific research. Part II: housing, husbandry and acquisition. *Animal Welfare, 15*(3), 215-238. doi:10.1017/S0962728600030451
Richig, J. W., & Sleeper, M. M. (2019). *Electrocardiography of Laboratory Animals*. London: Elsevier Inc.
Roelfsema, P. R., & Treue, S. (2014). Basic Neuroscience Research with Nonhuman Primates: A Small but Indispensable Component of Biomedical Research. *Neuron, 82*(6), 1200-1204. doi:10.1016/j.neuron.2014.06.003
Roma, P. G. (2006). The SPIT Method for Simultaneous and Unobtrusive Collection of Salivary Cortisol from Individually Housed Infant Monkeys. In G. P. Sackett, G. C. Ruppenthal, & K. Elias (Eds.), *Nursery rearing of nonhuman primates in the 21st century* (pp. 429-460). New York: Springer.
Rotenberg, S., & McGrath, J. J. (2016). Inter-relation between autonomic and HPA axis activity in children and adolescents. *Biological Psychology, 117*, 16-25. doi:10.1016/j.biopsycho.2016.01.015
Roughan, J. V., & Flecknell, P. A. (2006). Training in behaviour-based post-operative pain scoring in rats - An evaluation based on improved recognition of analgesic requirements. *Applied Animal Behaviour Science, 96*(3-4), 327-342. doi:10.1016/j.applanim.2005.06.012
Rushen, J., Chapinal, N., & de Passille, A. M. (2012). Automated monitoring of behavioural-based animal welfare indicators. *Animal Welfare, 21*(3), 339-350. doi:10.7120/09627286.21.3.339
Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*. London,: Methuen.
Sadoughi, B., Lacroix, L., Berbesque, C., Meunier, H., & Lehmann, J. (2021). Effects of social tolerance on stress: hair cortisol concentrations in the tolerant Tonkean macaques (Macaca tonkeana) and the despotic long-tailed macaques (Macaca fascicularis). *Stress-*

the International Journal on the Biology of Stress, 24(6), 1033-1041. doi:10.1080/10253890.2021.1998443

Salimetrics. (2021). Expanded Range High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit. Retrieved from https://salimetrics.com/wp-content/uploads/2018/03/salivary-cortisol-elisa-kit.pdf

Salimetrics. (2023). Saliva Collection Handbook. Retrieved from https://salimetrics.com/saliva-collection-handbook/

Salleh, M. R. (2008). Life event, stress and illness. *Malays J Med Sci, 15*(4), 9-18.

Sandi, C., & Pinelo-Nava, M. T. (2007). Stress and memory: behavioral effects and neurobiological mechanisms. *Neural Plast, 2007*, 78970. doi:10.1155/2007/78970

Schapiro, S. J., Bloomsmith, M. A., Kessel, A. L., & Shively, C. A. (1993). Effects of Enrichment and Housing on Cortisol Response in Juvenile Rhesus-Monkeys. *Applied Animal Behaviour Science, 37*(3), 251-263. doi:10.1016/0168-1591(93)90115-6

SCHEER, S. C. o. H., Environmental and Emerging Risks). (2017). Final Opinion on 'The need for non-human primates in biomedical research, production and testing of products and devices (update 2017)' *18 May 2017*.

Schmidt, A., Mostl, E., Wehnert, C., Aurich, J., Muller, J., & Aurich, C. (2010). Cortisol release and heart rate variability in horses during road transport. *Hormones and Behavior, 57*(2), 209-215. doi:10.1016/j.yhbeh.2009.11.003

Sellers, W. I., & Crompton, R. H. (2004). Automatic monitoring of primate locomotor behaviour using accelerometers. *Folia Primatol (Basel), 75*(4), 279-293. doi:10.1159/000078939

Sgoifo, A., Koolhaas, J. M., Musso, E., & De Boer, S. F. (1999). Different sympathovagal modulation of heart rate during social and nonsocial stress episodes in wild-type rats. *Physiology & Behavior, 67*(5), 733-738. doi:Doi 10.1016/S0031-9384(99)00134-1

Sha, J. C. M., Kaneko, A., Suda-Hashimoto, N., He, T. M., Take, M., Zhang, P., & Hanya, G. (2017). Estimating activity of Japanese macaques (Macaca fuscata) using accelerometers. *American Journal of Primatology, 79*(10). doi:10.1002/ajp.22694

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate variability Metrics and Norms. *Frontiers in Public Health, 5*. doi:UNSP 25810.3389/fpubh.017.00258

Sheriff, M. J., Dantzer, B., Delehanty, B., Palme, R., & Boonstra, R. (2011). Measuring stress in wildlife: techniques for quantifying glucocorticoids. *Oecologia, 166*(4), 869-887. doi:10.1007/s00442-011-1943-y

Sheriff, M. J., Krebs, C. J., & Boonstra, R. (2010). Assessing stress in animal populations: Do fecal and plasma glucocorticoids tell the same story? *General and Comparative Endocrinology, 166*(3), 614-619. doi:10.1016/j.ygcen.2009.12.017

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews, 68*, 651-668. doi:10.1016/j.neubiorev.2016.06.038

Shively, C. A., Appt, S. E., Chen, H. Y., Day, S. M., Frye, B. M., Shaltout, H. A., . . . Register, T. C. (2020). Mediterranean diet, stress resilience, and aging in nonhuman primates. *Neurobiology of Stress, 13*.

Shively, C. A., Mietus, J. E., Grant, K. A., Goldberger, A. L., Bennett, A. J., & Willard, S. L. (2007). Effects of chronic moderate alcohol consumption and novel environment on heart rate variability in primates (Macaca fascicularis). *Psychopharmacology, 192*(2), 183-191. doi:10.1007/s00213-007-0709-z

Shiverdecker, M. D., Schiml, P. A., & Hennessy, M. B. (2013). Human interaction moderates plasma cortisol and behavioral responses of dogs to shelter housing. *Physiology & Behavior, 109*, 75-79. doi:10.1016/j.physbeh.2012.12.002

Silverstein-Metzler, M. G., Frye, B. M., Justice, J. N., Clarkson, T. B., Appt, S. E., Carr, J. J., . . . Shively, C. A. (2022). Psychosocial stress increases risk for type 2 diabetes in female

cynomolgus macaques consuming a western diet. *Psychoneuroendocrinology, 139*. doi:ARTN 10570610.1016/j.psyneuen.2022.105706
Sinclair, R. C. F., & Faleiro, R. J. (2006). Delayed recovery of consciousness after anaesthesia. *Continuing Education in Anaesthesia Critical Care & Pain, 6*(3), 114-118. doi:10.1093/bjaceaccp/mkl020
Smith, G., D'Cruz, J. R., Rondeau, B., & Goldman, J. (2021). General Anesthesia for Surgeons. In *StatPearls*. Treasure Island (FL).
Smy, L., Shaw, K., Smith, A., Russell, E., Van Uum, S., Rieder, M., . . . Koren, G. (2015). Hair cortisol as a novel biomarker of HPA suppression by inhaled corticosteroids in children. *Pediatric Research, 78*(1), 44-47. doi:10.1038/pr.2015.60
Soltis, J., Wegner, F. H., & Newman, J. D. (2003). Adult cortisol response to immature offspring play in captive squirrel monkeys. *Physiology & Behavior, 80*(2-3), 217-223. doi:10.1016/j.physbeh.2003.07.009
Son, Y. (2010). Molecular mechanisms of general anesthesia. *Korean J Anesthesiol, 59*(1), 3-8. doi:10.4097/kjae.2010.59.1.3
Stalder, T., Steudte, S., Miller, R., Skoluda, N., Dettenborn, L., & Kirschbaum, C. (2012). Intraindividual stability of hair cortisol concentrations. *Psychoneuroendocrinology, 37*(5), 602-610. doi:10.1016/j.psyneuen.2011.08.007
Stanwicks, L. L., Hamel, A. F., & Novak, M. A. (2017). Rhesus macaques (Macaca mulatta) displaying self-injurious behavior show more sleep disruption than controls. *Applied Animal Behaviour Science, 197*, 62-67. doi:10.1016/j.applanim.2017.09.002
Steinmetz, H. W., Kaumanns, W., Dix, I., Heistermann, M., Fox, M., & Kaup, F. J. (2006). Coat condition, housing condition and measurement of faecal cortisol metabolites - a non-invasive study about alopecia in captive rhesus macaques (Macaca mulatta). *Journal of Medical Primatology, 35*(1), 3-11. doi:10.1111/j.1600-0684.2005.00141.x
Stephen, J. M., & Ledger, R. A. (2006). A longitudinal evaluation of urinary cortisol in kennelled dogs, Canis familiaris. *Physiology & Behavior, 87*(5), 911-916. doi:10.1016/j.physbeh.2006.02.015
Stilwell, G., de Carvalho, C., Lima, M. S., & Broom, D. M. (2008). The effect of duration of manual restraint during blood sampling on plasma cortisol levels in calves. *Animal Welfare, 17*(4), 383-385. doi:10.1017/S0962728600027883
Stonebarger, G. A., Bimonte-Nelson, H. A., & Urbanski, H. F. (2021). The Rhesus Macaque as a Translational Model for Neurodegeneration and Alzheimer's Disease. *Frontiers in Aging Neuroscience, 13*. doi:10.3389/fnagi.2021.734173
Stubsjoen, S. M., Knappe-Poindecker, M., Langbein, J., Fjeldaas, T., & Bohlin, J. (2015). Assessment of chronic stress in sheep (part II): Exploring heart rate variability as a non-invasive measure to evaluate cardiac regulation. *Small Ruminant Research, 133*, 30-35. doi:10.1016/j.smallrumres.2015.10.026
Stults-Kolehmainen, M. A., & Sinha, R. (2014). The Effects of Stress on Physical Activity and Exercise. *Sports Medicine, 44*(1), 81-121. doi:10.1007/s40279-013-0090-5
Sullivan, E. L., & Cameron, J. L. (2010). A rapidly occurring compensatory decrease in physical activity counteracts diet-induced weight loss in female monkeys. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology, 298*(4), R1068-R1074. doi:10.1152/ajpregu.00617.2009
Sullivan, E. L., Koegler, F. H., & Cameron, J. L. (2006). Individual differences in physical activity are closely associated with changes in body weight in adult female rhesus monkeys (Macaca mulatta). *American Journal of Physiology-Regulatory Integrative and Comparative Physiology, 291*(3), R633-R642. doi:10.1152/ajpregu.00069.2006
Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65*(1), 13-21. doi:10.1007/s00265-010-1037-6

Tannenbaum, J., & Bennett, B. T. (2015). Russell and Burch's 3Rs then and now: the need for clarity in definition and purpose. *J Am Assoc Lab Anim Sci, 54*(2), 120-132.
Tauson, R. (1998). Health and production in improved cage designs. *Poultry Science, 77*(12), 1820-1827. doi:10.1093/ps/77.12.1820
Thierman, J. L., Crowe, T. G., Stookey, J. M., & Valentine, B. (1999). Quantification of the response of elk during velvet antler removal. *Canadian Agricultural Engineering, 41*(4), 233-237. Retrieved from <Go to ISI>://WOS:000085503600004
Thierry, B. (2007). Unity in diversity: Lessons from macaque societies. *Evolutionary Anthropology, 16*(6), 224-238. doi:10.1002/evan.20147
Turini, L., Bonelli, F., Lanata, A., Vitale, V., Nocera, I., Sgorbini, M., & Mele, M. (2022). Validation of a new smart textiles biotechnology for heart rate variability monitoring in sheep. *Frontiers in Veterinary Science, 9*. doi:10.3389/fvets.2022.1018213
Uchiyama, H., Ohtani, N., & Ohta, M. (2007). The evaluation of autonomic nervous system activation during learning in rhesus macaques with the analysis of the heart rate variability. *Journal of Veterinary Medical Science, 69*(5), 521-526. doi:DOI 10.1292/jvms.69.521
Ueda, Y., Slabaugh, T. L., Walker, A. L., Ontiveros, E. S., Sosa, P. M., Reader, R., . . . Stern, J. A. (2019). Heart Rate and Heart Rate Variability of Rhesus Macaques (Macaca mulatta) Affected by Left Ventricular Hypertrophy. *Frontiers in Veterinary Science, 6*. doi:10.3389/fvets.2019.00001
Unakafov, A. M., Möller, S., Kagan, I., Gail, A., Treue, S., & Wolf, F. (2018). Using imaging photoplethysmography for heart rate estimation in non-human primates. *Plos One, 13*(8), e0202581. doi:10.1371/journal.pone.0202581
Understanding Animal Research. (2007). Medical Advances and Animal Research. The Contribution of Animal Science to the Medical Revolution: Some Case Histories. *RDS: Understanding Animal Research in Medicine and Coalition for Medical Progress*.
Understanding Animal Research. (2014). *Concordat on Openness on Animal Research in the UK*. https://concordatopenness.org.uk/.
van der Meij, E., van der Ploeg, H. P., van den Heuvel, B., Dwars, B. J., Meijerink, W. J. H. J., Bonjer, H. J., . . . Anema, J. R. (2017). Assessing pre- and postoperative activity levels with an accelerometer: a proof of concept study. *BMC Surgery, 17*(1). doi:10.1186/s12893-017-0223-0
Vandeleest, J. J., Capitanio, J. P., Hamel, A., Meyer, J., Novak, M., Mendoza, S. P., & McCowan, B. (2019). Social stability influences the association between adrenal responsiveness and hair cortisol concentrations in rhesus macaques. *Psychoneuroendocrinology, 100*, 164-171. doi:10.1016/j.psyneuen.2018.10.008
Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., . . . Würbel, H. (2020). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience, 21*(7), 384-393. doi:10.1038/s41583-020-0313-3
von Borell, E., Langbein, J., Despres, G., Hansen, S., Leterrier, C., Marchant-Forde, J., . . . Veissier, I. (2007). Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - A review. *Physiology & Behavior, 92*(3), 293-316. doi:10.1016/j.physbeh.2007.01.007
von der Ohe, C. G., & Servheen, C. (2002). Measuring stress in mammals using fecal glucocorticoids: opportunities and challenges. *Wildlife Society Bulletin, 30*(4), 1215-1225. Retrieved from <Go to ISI>://WOS:000180540900028
Vutskits, L., & Xie, Z. C. (2016). Lasting impact of general anaesthesia on the brain: mechanisms and relevance. *Nature Reviews Neuroscience, 17*(11), 705-717. doi:10.1038/nrn.2016.128
Walther, A. R., Stepula, E., Ditzel, N., Kassem, M., Bergholt, M. S., & Hedegaard, M. A. B. (2023). In Vivo Longitudinal Monitoring of Disease Progression in Inflammatory Arthritis Animal

Models Using Raman Spectroscopy. *Analytical chemistry, 95*(7), 3720-3728. doi:10.1021/acs.analchem.2c04743
Wasowicz-Kemps, D. K., Slootmaker, S. M., Kemps, H. M. C., Borel-Rinkes, I. H. M., Biesma, D. H., & Van Ramshorst, B. (2009). Resumption of daily physical activity after day-case laparoscopic cholecystectomy. *Surgical Endoscopy, 23*(9), 2034-2040. doi:10.1007/s00464-008-9928-6
Watanabe, N., Sakanoue, S., Kawamura, K., & Kozakai, T. (2008). Development of an automatic classification system for eating, ruminating and resting behaviour of cattle using an accelerometer. *Japanese Society of Grassland Science*.
Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., & Naito, Y. (2005). A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. *Applied Animal Behaviour Science, 94*(1-2), 117-131. doi:10.1016/j.applanim.2005.01.010
Watters, J. V., Krebs, B. L., & Eschmann, C. L. (2021). Assessing Animal Welfare with Behavior: Onward with Caution. *Journal of Zoological and Botanical Gardens, 2*(1), 75-87. doi:10.3390/jzbg2010006
Weatherall, D., Goodfellow, P, Harris, J, Hinde, R, Johnson, L, Morris, R, Ross, N, Skehel, J and Tickell, C. (2006). The use of non-human primates in research. *London: Academy of Medical Sciences.* .
Webster, J. (2016). Animal Welfare: Freedoms, Dominions and "A Life Worth Living". *Animals (Basel), 6*(6). doi:10.3390/ani6060035
Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the Anatomy, Physiology, and Pharmacology of the Autonomic Nervous System. *Comprehensive Physiology, 6*(3), 1239-1278. doi:10.1002/cphy.c150037
Whitten, P. L., Stavisky, R., Aureli, F., & Russell, E. (1998). Response of fecal cortisol to stress in captive chimpanzees (Pan troglodytes). *American Journal of Primatology, 44*(1), 57-69. doi:Doi 10.1002/(Sici)1098-2345(1998)44:1
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org
Williams, B. (2020). *Public attitudes to animal research under COVID-19*. Understanding Animal Research Retrieved from https://www.understandinganimalresearch.org.uk/files/3315/8687/3612/attitudes_to_animal_research_under_covid-19_final.pdf
Wilson, A. M., & Lipworth, B. J. (1999). Short-term dose-response relationships for the relative systemic effects of oral prednisolone and inhaled fluticasone in asthmatic adults. *Br J Clin Pharmacol, 48*(4), 579-585. doi:10.1046/j.1365-2125.1999.00058.x
Wolf, M., & Weissing, F. J. (2012). Animal personalities: consequences for ecology and evolution. *Trends in Ecology & Evolution, 27*(8), 452-461. doi:10.1016/j.tree.2012.05.001
Wood, E. K., Halter, C. M., Gabrielle, N., Capitanio, J. P., & Higley, J. D. (2021). Stress-induced plasma cortisol concentrations in infancy are associated with later parenting behaviors in female rhesus macaques (Macaca mulatta). *Developmental Psychobiology, 63*(5), 1098-1108. doi:10.1002/dev.22098
Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statisticalk Society (B), 73*(1), 3-36.
Wright-Williams, S. L., Courade, J. P., Richardson, C. A., Roughan, J. V., & Flecknell, P. A. (2007). Effects of vasectomy surgery and meloxicam treatment on faecal corticosterone levels and behaviour in two strains of laboratory mouse. *Pain, 130*(1-2), 108-118. doi:10.1016/j.pain.2006.11.003
Wunderlich, R. E., Tongen, A., Gardiner, J., Miller, C. E., & Schmitt, D. (2014). Dynamics of Locomotor Transitions from Arboreal to Terrestrial Substrates in Verreaux's Sifaka

(Propithecus verreauxi). *Integrative and Comparative Biology, 54*(6), 1148-1158. doi:10.1093/icb/icu110

Yamanashi, Y., Morimura, N., Mori, Y., Hayashi, M., & Suzuki, J. (2013). Cortisol analysis of hair of captive chimpanzees (Pan troglodytes). *Gen Comp Endocrinol, 194*, 55-63. doi:10.1016/j.ygcen.2013.08.013

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482.

Yoshitoshi, R., Watanabe, N., Kawamura, K., Sakanoue, S., Mizoguchi, R., Lee, H. J., & Kurokawa, Y. (2013). Distinguishing Cattle Foraging Activities Using an Accelerometry-Based Activity Monitor. *Rangeland Ecology & Management, 66*(3), 382-386. doi:10.2111/Rem-D-11-00027.1

Young, C., Majolo, B., Heistermann, M., Schulke, O., & Ostner, J. (2014). Responses to social and environmental stress are attenuated by strong male bonds in wild macaques. *Proceedings of the National Academy of Sciences of the United States of America, 111*(51), 18195-18200. doi:10.1073/pnas.1411450111

Young, K. M., Walker, S. L., Lanthier, C., Waddell, W. T., Monfort, S. L., & Brown, J. L. (2004). Noninvasive monitoring of adrenocortical activity in carnivores by fecal glucocorticoid analyses. *Gen Comp Endocrinol, 137*(2), 148-165. doi:10.1016/j.ygcen.2004.02.016

Yuda, E., Shibata, M., Ogata, Y., Ueda, N., Yambe, T., Yoshizawa, M., & Hayano, J. (2020). Pulse rate variability: a new biomarker, not a surrogate for heart rate variability. *Journal of Physiological Anthropology, 39*(1). doi:10.1186/s40101-020-00233-x

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliey, A. A., & Smith, G. M. (2009). *Mixed Effect Models and Extensions in Ecology with R*. New York: Springer-Science.

Appendices

Appendix 1

The impact of general anaesthesia on non-human primates

Bootstrap analyses

To determine the degree of variability in baseline daytime activity values of rhesus macaques, mean daily activity values (12 hours) were calculated and plotted for a series of candidate time windows for data prior to interventions, using two alternative randomisation methods. Randomisations were applied separately to each event. *Method 1* involved randomly selecting, with replacement, daily activity values from the pool of 'prior' data for sample sizes ranging from 1 to 10 days (Figure A1a). *Method 2* also involved randomly sampling with replacement from the pool of available data. However, this method also included an autocorrelation structure to account for inherent dependent present in macaque activity between days. Here, data within sample draws for a particular time window were constrained to be taken from a sequential block in time (Figure A1b).

The standard deviation (SD) of daytime activity values decreased steadily with an increasing baseline time window. For both randomisation methods, knee-points in SD values occurred at time windows of 2 or 3 days (Figure A1). Therefore, only used events with 3 or more data points for data prior to interventions.

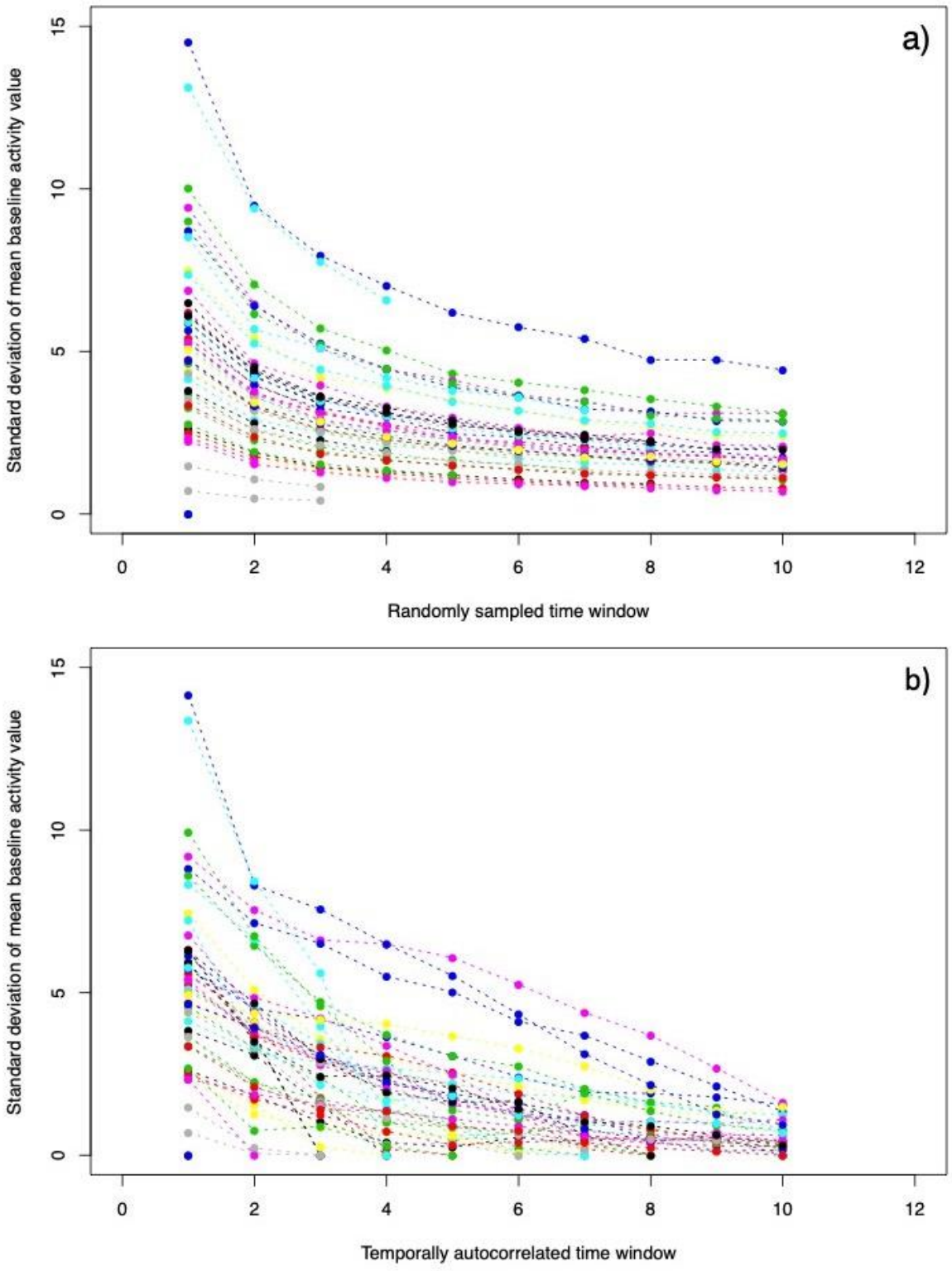

Figure A1: The variability (standard deviation) around mean daytime activity values of macaques prior to MRI interventions for a) randomly sampled and b) temporally auto-correlated randomly sampled draws across increasing time windows (days). Dotted lines join data from the same event across time windows.

Piecewise linear regression

A series of candidate piecewise regression models were run during data exploratory phases to investigate the time taken to return to baseline activity levels following an MRI (using methods in (Berman et al., 2009)). All models included a random effects term for individual to account for non-independence of observations from the same animal, and 'days since MRI intervention' as a fixed effect. Models had a single threshold, the slopes either side of which were allowed to vary. Models with thresholds ranging from 2 to 14 days were compared using Akaike Information Criterion (AIC), and the threshold that minimised the AIC (where ΔAIC >2 between models; (Burnham & Andersen, 2002)) was taken to the most parsimonious. Models with different random effects structures were also compared using restricted maximum likelihood estimated (REML) to investigate the effect of allowing both the intercept and slopes either side of the best threshold to vary with individual.

Macaques exhibited a significant response in daytime activity to MRI interventions with the best threshold model explaining more variation in the data than a null model that included just an individual random effect (LRT, R^2= 74.741, *P* <0.001; Table A1). The best model was that with a threshold term of 4 days, representing an estimated time to return to baseline activity levels at the population level (Figure A2; Table A2). Addition of a random slope for individual did not improve the model (ΔAIC of 'intercept only' model compared to 'Intercept and slope' models *with correlation* = +2.734 and *without correlation* = +0.887), and the random individual intercept explained only a small amount of variance in the data (variance of random intercept = 1.975, residual variance = 59.837).

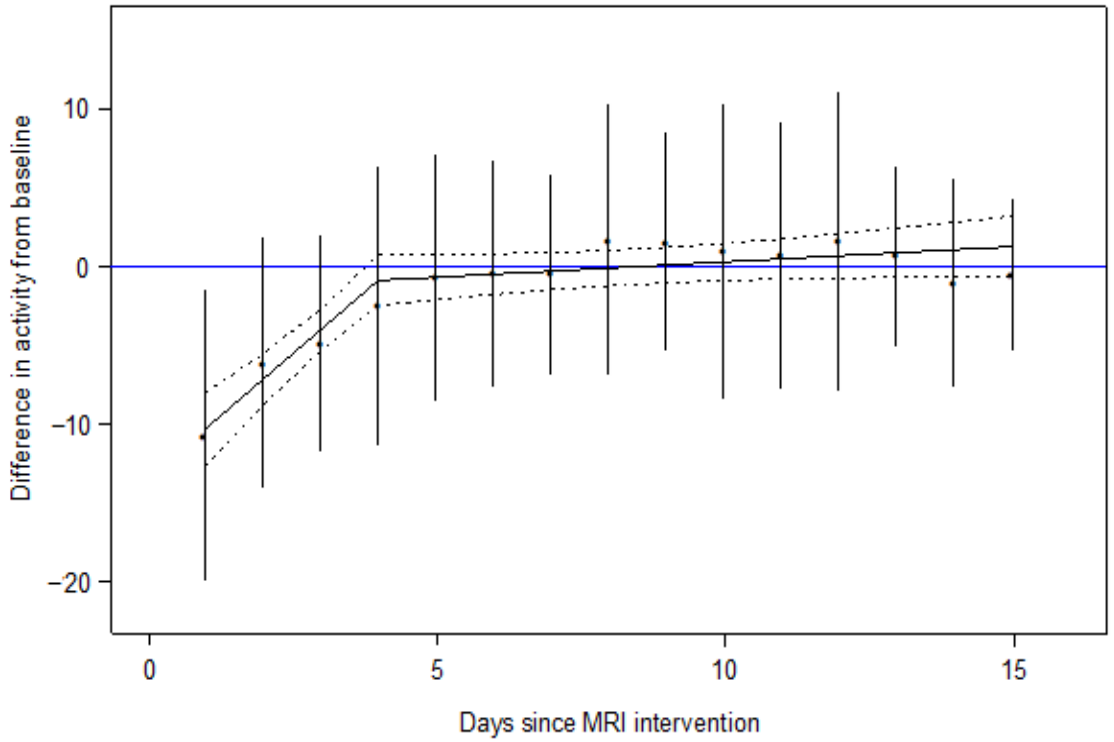

Figure A2: Fitted values (black solid line) and 95% confidence bands (dotted lines) from a single threshold piecewise linear mixed-effects model (threshold = 4 days) run to investigate the relationship between days since MRI intervention and difference in activity from baseline levels in macaques (n, individual = 23, intervention = 35). Blue solid line = activity difference of zero. Points and vertical bars = mean ± SD values in daytime activity.

Table A1: Results of model selection for piecewise linear mixed-effects models (threshold = 4 days) with a random 'individual' intercept (Ind), applied to test the relationship between days since MRI intervention (Delta) and the difference in daytime activity from baseline levels (ActDiff) in macaques (n, individual = 23, MRI intervention = 35). Measures of Akaike Information Criteria (AIC), delta AIC components and Akaike weights are shown. The most parsimonious model is shown in bold.

Model	Meaning	AIC	ΔAIC	AIC wt
ActDiff ~ 1 + (1|Ind)	No effect of Delta on ActDiff	3134.0	68.1	<0.001
ActDiff ~ Delta + (1|Ind)	Linear effect of Delta on ActDiff	3093.4	27.6	<0.001
ActDiff ~ DT_1 + DT_2 + (1|Ind)	**Single threshold Delta**	**3065.9**	**0.0**	**0.999**

Table A2: Parameter estimates from the most parsimonious pairwise linear mixed-effects model (threshold = 4 days; random intercept for individual), run to investigate the effect of days since MRI intervention on difference in activity from baseline levels in macaques (n, individual = 23, MRI intervention = 35; df = 424).

Term	Parameter estimate	Standard error	*t*-value
Intercept	13.472	1.565	8.608
Slope before threshold	-3.146	0.467	-6.744
Slope after threshold	-0.195	0.123	-1.584
		Random intercept	**Residual**
	Variance	1.238	59.732

Appendix 2

Automated behavioural assessment using accelerometers

The model was created by exploring corresponding actigraphy data and behavioural observations to identify ranges and patterns in the activity levels of each behaviour. Initially, the rule-based model and it's 5 behavioural categories (see Chapter 4 - Table 4.1) were identified during a pilot study and appeared to predict behaviours with a good overall accuracy (80%). However, the pilot study dataset consisted of observations from only one observer and when new observers scored the same data, the accuracy of that model dropped significantly (~60%). Following on from this, intra and inter-rater variability were assessed and the rules of the model were revisited, using either the mode of behavioural observations when multiple scores were available or the score of RH (myself) who was the most experienced and consistent (intra-rater variability) observer.

Several attempts were made to adjust the pilot study ruleset to accurately predict the new behavioural observations, but accuracy remained low (~60%). To that end, a new set of rules for the model was created. This began with mapping out the typical raw activity levels that fell within each behavioural category and identifying overlap between behaviours. There appeared to be significant overlap between some behaviours and therefore the decision was made to broaden the initial behavioural categories such that some behaviours were now grouped together (see Table A3). To determine what rules could be applied to then split these grouped categories into the 5 targeted behaviours, activity epochs from each behaviour were investigated to establish patterns. For example, it was observed that pacing behaviours typically lasted longer than 30 seconds (3 x 10 second epochs) and the standard deviation between the epochs within a bout was quite low, as pacing was a consistent pattern of repetitive locomotion. The work to identify patterns such as this helped to establish the final 6 rules. Once the rules were established, behavioural observations that were not used to create the rules (i.e. a rule refinement dataset)

were compared to model predictions and then numerous adjustments were made to the rule thresholds in order to improve accuracy. Furthermore, observations from new animals (i.e. those not used to create the rules) were also used to refine the model, to ensure the model was not overly fitted to a few individuals. Typically, 1 hour of footage at a time was used for rule refinement and several iterations of the rule thresholds were tried to create the final model.

Knowledge-based model rules

Step 1: Assign each epoch 1 out of 7 behavioural categories based on raw activity value (Table A3).

Table A3: Behavioural categorisation based on raw activity value alone. Step 1 of the knowledge-based model.

Activity value	Behavioural category	Behaviours represented
0	**IAAR**	Inactive & Active rest
1-24	**AR**	Active rest
25-79	**GL**	General Locomotion
80-89	**GLP**	General Locomotion & Pacing
90-99	**GLPR**	General Locomotion & Rapid Movement & Pacing
100-169	**PR**	Pacing & Rapid Movement
170+	**R**	Rapid Movement

Step 2: Separate **IAAR** category into either Inactive (**IA**) or Active rest (**AR**) behaviours.

- If 4 or more consecutive epochs are **IAAR**, change all epochs to **IA**.
- If less than 4 consecutive epochs are **IAAR**, change all epochs to **AR**.

Step 3: Define bouts of pacing behaviour.

- Any epoch assigned a category that represents a pacing behaviour (**GL*P*, GL*P*R, *P*R**) is considered a 'pacing epoch.'
- 'Pacing bouts' are lengths of consecutive 'pacing epochs' uninterrupted by:
    - any **IA** or **AR** epochs.
    - > 2 epochs of **GL** or **R**.

Step 4: Redefining pacing bouts.

- Mean, standard deviation, and length of current 'pacing bouts' is calculated.
- 'Pacing bouts' are declassified as such if:
    - Length < 3 epochs.
    - Length < 5 epochs AND standard deviation is > 35.
- Assign epochs within 'pacing bouts' to a behavioural category.
    - All **GLPR, GLP, PR** and **GL** behaviours within a bout change to **P**.
    - All **R** behaviours within a bout remain **R**.

Step 5: Remove pacing behaviour outside of pacing bouts.

- Epochs that do not sit within pacing bouts that contain '**P**' behaviour should remove **P**.
    - **GLP** and **PR** change to **GL** and **R** respectively.
    - **GLPR** changes to **GLR**.

Step 6: Separate remaining **GLR** epochs.

- If the difference between **GLR** epoch and the preceding epoch >= 70, change epoch to **R**.
- If the difference between **GLR** epoch and the preceding epoch < 70, change epoch to **GL**.

Appendix 3

Cortisol and non-human primate welfare

Table A4: Outputs from the linear-mixed model looking at influence of multiple fixed effects on the log of hair cortisol concentration (all data from 22 male macaques, except for three outliers, were used in the model). Significant effects can be seen in bold.

Fixed Effects	*Estimate*	*SE*	*t*	*d.f.*	*p*
Intercept	2.23	0.30	7.37	46.48	
Time In Oxford (years)	0.02	0.02	1.28	354.22	0.20
Age On Arrival (years)	0.13	0.07	2.00	30.71	0.05
Single Housing (yes/no)	0.12	0.09	1.26	12.52	0.23
GA/MRI (yes/no)	0.08	0.05	1.63	507.33	0.10
Surgery (yes/no)	**0.22**	**0.05**	**4.74**	**507.98**	**<0.001**

Pseudo-R^2 (fixed effects) = 0.10; Pseudo-R^2 (total) = 0.53; ICC (Identity) = 0.37; ICC (Social Group) = 0.11.

Table A5: Linear-mixed model output looking at the influence of the time spent in storage (days) on the log of hair cortisol concentration (all data from 22 male macaques, except for three outliers, were used in the model). Significant effects can be seen in bold.

Fixed Effects	*Estimate*	*SE*	*t*	*d.f.*	*p*
Intercept	3.00	0.08	38.16	57.67	
Length of Storage (days)	**-0.0001**	**0.00004**	**-2.37**	**375.52**	**0.02**

Pseudo-R^2 (fixed effects) = 0.009; Pseudo-R^2 (total) = 0.48; ICC (Identity) = 0.37; ICC (Social Group) = 0.11

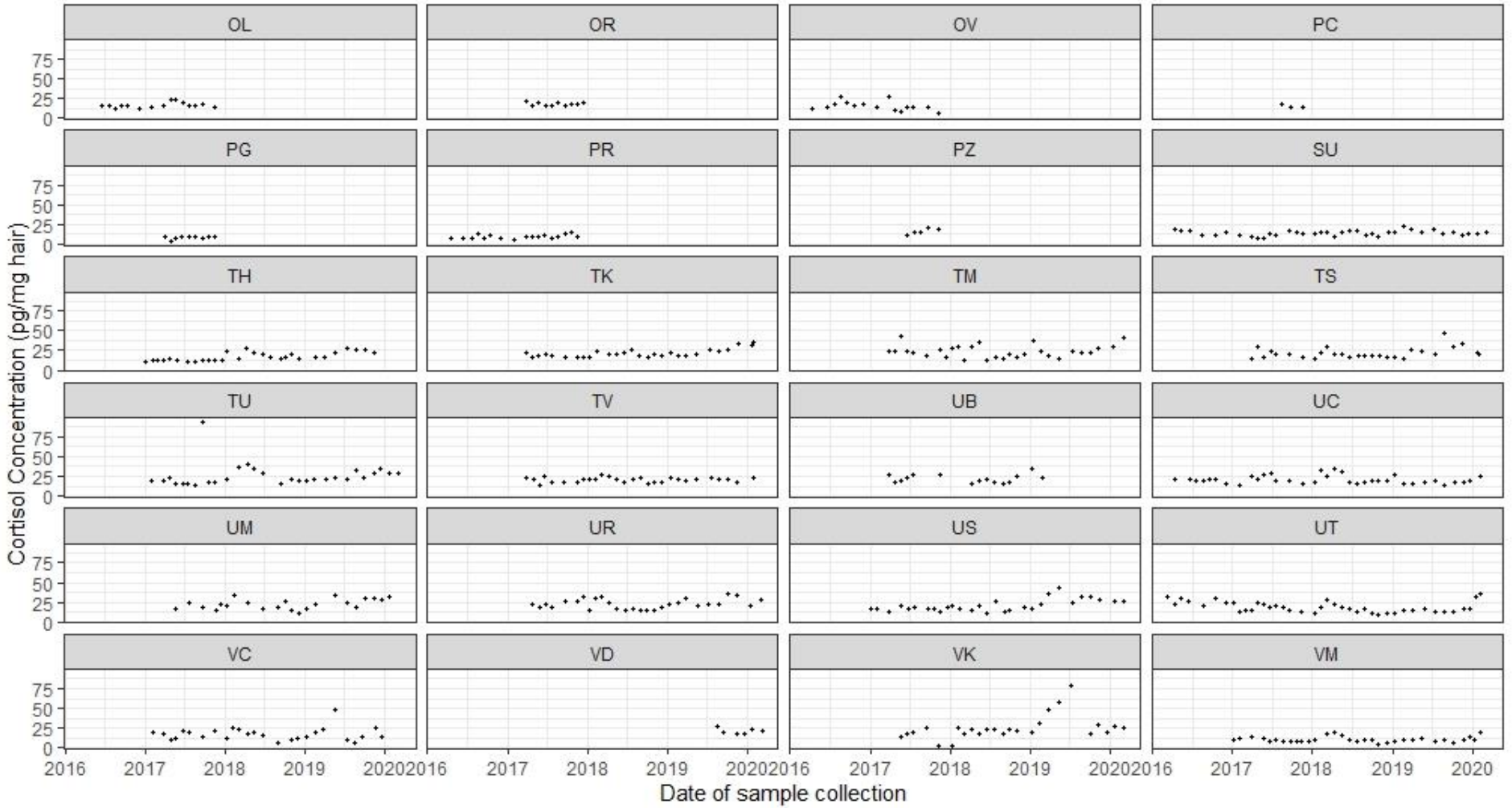

Figure A3: Hair cortisol concentration of 24 male macaques over time.