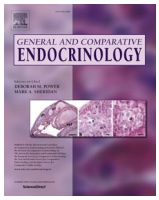
General and Comparative Endocrinology 342 (2023) 114351

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen



Research paper

An empirical comparison of several commercial enzyme immunoassays for the non-invasive assessment of adrenocortical and gonadal function in mountain gorillas

Nicholas M. Grebe [a,*], Winnie Eckardt [b], Tara S. Stoinski [c], Rose Umuhoza [b], Rachel M. Santymire [d], Stacy Rosenbaum [a]

[a] Department of Anthropology, University of Michigan, Ann Arbor MI, United States
[b] Dian Fossey Gorilla Fund, Musanze, Rwanda
[c] Dian Fossey Gorilla Fund, Atlanta GA, United States
[d] Department of Biology, Georgia State University, Atlanta GA, United States

ARTICLE INFO

Keywords:
Wildlife
Primate
Immunoassay
Assay validation
Androgen
Glucocorticoid

ABSTRACT

Wildlife researchers seeking to non-invasively examine endocrine function in their study species are presented with a dense and technical 'garden of forking paths' to navigate between collecting a biological sample and obtaining a final measurement. In particular, the choice of which enzyme immunoassay (EIA) to use with collected fecal samples, out of the many options offered by different manufacturers and research laboratories, may be one of the most consequential for final results. However, guidance for making this decision is still emerging. With this gap in mind, we performed a head-to-head comparison of results obtained from four different EIAs for fecal glucocorticoid metabolites (FGCMs), and three different EIAs for fecal androgen metabolites (FAMs), applied to the same set of fecal samples collected from the mountain gorillas (*Gorilla beringei beringei*) monitored by the Dian Fossey Gorilla Fund in Volcanoes National Park, Rwanda. We provide a) an analytical validation of the different EIAs via tests of parallelism and linearity; b) an estimate of inter-assay correlation between EIA kits designed for the same metabolites; and c) a test of the kits' ecological validity, in which we examine how well each captures endocrine changes following events that theory predicts should result in elevated FGCM and/or FAM concentrations. Our results show that kits differ to some degree in their performance; at the same time, nearly all assays exhibited at least moderate evidence of validity and covariance with others for the same analyte. Our findings, which differ somewhat from similar comparisons performed in other species, demonstrate the need to directly assess assay performance in a species- and context-specific manner as part of efforts to develop the burgeoning discipline of wildlife endocrinology.

1. Introduction

Wildlife researchers investigate a wide variety of questions and taxa, but they share a common objective of monitoring and preserving the health and well-being of their study species. One important way to examine these broad constructs involves 'getting under the skin' of individuals to study aspects of their physiology, such as reproductive or adrenocortical functioning (Schwarzenberger, 2007; Romano et al., 2010; Ganswindt et al., 2012; Behringer & Deschner, 2017; King et al., 2023). Measuring these physiological biomarkers can reveal and/or confirm a variety of dynamics relevant to behavioral ecology and wildlife conservation, including the transmission of communicable diseases, social predictors of physiological states, anthropogenic impacts on energetic status and stress physiology, and variation in reproductive/breeding potential (Touma & Palme, 2005; Eckardt et al., 2016). In captive settings, physiological parameters can often be directly assessed by drawing blood or obtaining other samples or measurements that require close contact, but the need for non-invasive sampling in wildlife research has driven substantial interest in methods for assaying biomarkers through media, such as feces, urine, or hair. The measurement of fecal steroid hormone metabolites in particular forms the basis for a wide variety of non-invasive research in wildlife studies, veterinary

* Corresponding author.
E-mail address: ngrebe@umich.edu (N.M. Grebe).

https://doi.org/10.1016/j.ygcen.2023.114351
Received 20 February 2023; Received in revised form 29 June 2023; Accepted 27 July 2023
Available online 31 July 2023

medicine, and conservation biology (see Palme, 2005; Touma & Palme, 2005; Wielebnowski & Watters, 2007; Sheriff et al., 2011).

Researchers seeking to measure a particular biomarker from fecal samples are presented with a dizzying 'garden of forking paths' (Gelman & Loken, 2014) for processing and analysis (see e.g. Palme, 2005; Wielebnowski & Watters, 2007)–one that is multiplied for each additional biomarker. At numerous junctures along the path from 'raw sample' to 'final measurement', there exist multiple options for proceeding forward, and often, even well-established methods vary in their performance. For instance, while there is wide agreement that raw fecal samples must undergo an extraction process prior to assay, the use of different extraction methods (e.g. varieties of filter columns or solvents) on the same samples may yield different final measurements, even if all subsequent analytical steps are identical (Beehner and Whitten, 2004; Pappano et al., 2010; Palme et al., 2013). Researchers must next select a quantification method to use on their sample extract, with mass spectrometry and immunoassay-based methods being the most common classes of technique. Previous studies suggest that mass spectrometry and radioimmunoassay methods may possess advantages such as increased specificity and precision of measurement (see e.g. Murtagh et al., 2013; Welker et al., 2016; Arslan et al., 2023). However, wildlife researchers constrained by costs and logistics have increasingly moved toward enzyme immunoassays (EIAs; Palme, 2005; Touma & Palme, 2005; Wielebnowski & Watters, 2007). EIAs are relatively cost-effective and, unlike radioimmunoassays, do not rely on the use of radioactive materials; consequently, they have become the typical method of choice for fecal biomarker assays.

Even though the popularity and acceptance of EIAs to measure fecal hormone metabolites has rapidly increased within the field of wildlife biology, and a number of methodological reviews have attempted to establish best practices for measurement (e.g. Palme, 2005; Touma & Palme, 2005; Wielebnowski & Watters, 2007; Palme et al., 2013; Palme, 2019), researchers are still tasked with choosing between a number of defensible options when validating an assay for a biomarker of interest. Most prominently, the assays advertised to measure fecal metabolites typically use different proprietary detection antibodies, which might be designed to target the same analyte (e.g. cortisol) or different components of the same hormonal axis (e.g. 11-oxoetiocholanolone metabolites vs. corticosterone; Hinchcliffe et al., 2021). Previous research has established that antibody selection influences measurement of hormone metabolites (Touma & Palme, 2005; Hinchcliffe et al., 2021), making choice of antibody one of the most consequential for any study that relies on estimation of fecal hormone metabolites.

In this paper, we explore the extent to which several different EIA options commonly employed by wildlife biologists converge on similar answers. In particular, using the same sets of mountain gorilla fecal samples, we provide a head-to-head comparison of results from multiple assays marketed to assess fecal glucocorticoid metabolites (FGCMs) and fecal androgen metabolites (FAMs). In our experience, researchers often choose among these options based on inertia, convenience, and cost, and there is an implicit assumption that all assays for FGCMs or FAMs will roughly yield the same results. However, given that these assays (a) use different antibodies and reagents; (b) may be targeted to detect distinct analytes; and (c) may function differently depending on the study species, this assumption requires rigorous testing (as has been argued extensively elsewhere, e.g. Touma & Palme, 2005; Wielebnowski & Watters, 2007; Palme, 2019).

Comparing assay methods to one another does not provide guidance on which assays appropriately reflect an organism's internal physiological processes. To address this question, biological validation is necessary (Touma & Palme, 2005). Traditional physiological validations that experimentally modify upstream endocrine pathways to generate reliable changes in hormone concentrations–such as an adrenocorticotropic hormone (ACTH) challenge or a dexamethasone suppression test for HPA functioning, or a gonadotropin-releasing hormone (GnRH) challenge for HPG functioning–are not possible for mountain gorillas, who are endangered and exist only in the wild. Thus, we rely on two previously published biological validations of fecal steroid metabolites as a benchmark. In these validation studies, samples were collected before and after competitive events. One study focused on between-group interactions, which was an event hypothesized and then confirmed to upregulate HPA axis activity, as expressed via marked increases in FGCMs (Eckardt et al., 2016). A second study assessed a broader range of competitive events and similarly found increases in FAMs in males following these events (Rosenbaum et al., 2021).

Notably, the detection antibodies used in Eckardt et al. (2016) and Rosenbaum et al. (2021) were manufactured in an academic lab. They are no longer readily available for many researchers, and the remaining supply will eventually disappear completely. We anticipate challenges to the continuing availability of assay materials will be encountered by many labs and field sites engaged in long-term monitoring of wildlife, making a comparison between validated older assays and newer, more widely available commercial assays timely for both analytical and practical purposes. We provide such a comparison analyzing the same samples assayed in Eckardt et al. (2016) and Rosenbaum et al. (2021) with multiple commercially available immunoassays. Specifically, for FGCMs, we compare the older assay used in Eckardt et al. (2016) with three newer commercial immunoassays, including two developed specifically for wildlife researchers by the International Society of Wildlife Endocrinology (ISWE). We specifically sought to examine performance of the ISWE assays relative to other offerings due to ISWE's stated goal to provide cost-effective research tools for the non-invasive measurement of hormones and hormone metabolites in wildlife (see e.g. Ganswindt et al., 2012). For FAMs, in a dataset of male samples, we compare the assay used in Rosenbaum et al. (2021) with two commercial EIAs for testosterone, one of which was also developed by ISWE. Thus, in addition to providing an analytical, internal validation of these EIAs for mountain gorilla fecal samples, we also provide an external validation by testing these kits' ability to capture natural hormone increases.

2. Methods

The mountain gorillas whose fecal samples were used for this study lived in 10 social units in Volcanoes National Park, Rwanda. These groups are continually monitored by the Dian Fossey Gorilla Fund. In the Virunga massif, which contains Volcanoes National Park, elevation ranges from 2300 to 4500 m altitude and temperatures are mild year-round. Rainfall is bimodally distributed, with a long and short wet season lasting from September to December and March to May, respectively (Eckardt et al., 2019).

2.1. Sample collection and processing

Our overall dataset stems from samples that were initially analyzed in Eckardt et al. (2016) and Rosenbaum et al. (2021). Samples used in these papers were collected from April 2011 - December 2012. Jointly, these papers report on a total of 134 fecal samples that were assayed–all 134 for FGCMs, with 48 samples from adult males additionally assayed for FAMs. Six samples from adult males in Eckardt et al. (2016) were not assayed for FAMs in Rosenbaum et al. (2021), but we included these samples in our two newer FAM assays to maximize sample size ($n = 54$ for these assays).

Details of collection are reported extensively in Eckardt et al. (2016) and Rosenbaum et al. (2021). Briefly, between 0700 h and 1600 h on a given observation day, trained observers able to identify individual animals collected fecal samples uncontaminated with rainwater, urine, or other gorillas' fecal material. Samples were placed in a cooler bag with an ice pack at the end of the day of observation, and upon return to the KRC lab the same day, the samples were placed in a −20 °C freezer. A trained KRC laboratory technician (author RU) completed the extractions for all samples within 2–3 months after their initial freeze, using a validated, field-friendly technique that has previously been applied to

other samples from this gorilla population (Eckardt et al., 2016; Santymire & Armstrong, 2010). After extraction, samples were dried down, capped, and returned to the −20 °C freezer until shipment to US-based laboratories. All evaporated extracts were kept frozen at −20 °C until being reconstituted in assay buffer and, if necessary, diluted for laboratory assays.

Biological Validation. Extensive details of the biological validations are also provided in Eckardt et al. (2016) and Rosenbaum et al. (2021). Briefly, the authors collected weekly "baseline" samples from each animal, as well as all possible samples in the 6 days following a designated event expected to relate to changes in glucocorticoid and/or androgen secretion: for FGCMs, the focal events used in the publication were all intergroup interactions; for FAMs, sampling was targeted at males following an intragroup fight, an intergroup fight, a female transfer, and leadership of a temporary subgroup.

2.2. In-House FGCM and FAM assays

The original biological validation assay for FGCMs presented in Eckardt et al. (2016), and the original biological validation assay for FAMs presented in Rosenbaum et al. (2021), were both performed at the Davee Center for Epidemiology and Endocrinology at Lincoln Park Zoo (Chicago, IL, USA) by authors SR, RS, and laboratory technicians. FGCMs and FAMs were quantified using enzyme immunoassays with an in-house detection antibody for either cortisol (R4866; 1:20000 dilution; hereafter identified as "R4866 Cortisol") or testosterone (R156/7; 1:10000; "R156/7 Testosterone"), both of which were provided by C.J. Munro, University of California, Davis. These assays were carried out according to previously published procedures (e.g., Loeding et al., 2011; Murray et al., 2013; Freeman et al., 2014). Samples were first assayed at a 1:5 dilution (this ratio refers to feces wet mass: ethanol volume during the initial extraction; extracts were then reconstituted in an equivalent amount of assay buffer) for R4866 Cortisol, and 1:250 for R156/7 testosterone, and were re-run with higher or lower dilutions as necessary when measured concentrations fell above or below the range, respectively, of the standard curve. All concentrations (ng/g wet feces) reported in results are corrected for dilution factor.

2.3. Commercial FGCM and FAM assays

All additional assays were performed at the University of Michigan Primate Behavior Lab by authors NG and SR. Using the same samples assayed in Eckardt et al. (2016) and Rosenbaum et al. (2021), we created new dilutions from the stock 1:5 extractions for use in five additional assays–three for FGCMs, and two for FAMs–all of which were distributed by Arbor Assays (Ann Arbor, MI) but were developed for distinct purposes. Within hormone metabolite group (FGCMs or FAMs), the most relevant difference between assays concerns the primary antibody used. For FGCMs, we compared 1) a popular commercial cortisol EIA using a monoclonal antibody (hereafter "Arbor Assays Cortisol"; Catalog #K-003H), 2) an EIA with a polyclonal antibody developed by ISWE to detect both cortisol and its excreted metabolites ("ISWE Cortisol"; Catalog #ISWE-002), and 3) an EIA with a polyclonal antibody developed by ISWE to detect both corticosterone and its excreted metabolites ("ISWE Corticosterone"; Catalog #ISWE-007). For FAMs, we compared 1) a popular commercial testosterone EIA using a polyclonal antibody ("Arbor Assays Testosterone"; Catalog #K-032H) and 2) an EIA with a different polyclonal antibody developed by ISWE to detect both testosterone and its excreted metabolites ("ISWE Testosterone"; Catalog #ISWE-001). While the antibodies developed by Arbor Assays have had some success in detecting hormone metabolites in past primate research (e.g., Pollastri et al., 2022), we examined the possibility that ISWE antibodies specifically developed for wildlife research are more suitable to detect a broader range of hormone metabolites. Samples were first run at a 1:10 dilution for both cortisol assays, 1:40 for ISWE Corticosterone, and 1:80 for both testosterone assays. We re-ran samples at higher or lower dilutions as necessary when measured concentrations fell above or below the range, respectively, of the standard curve. All concentrations (ng/g wet feces) reported in results are corrected for dilution factor.

For each assay, we calculated the overall intraassay coefficient of variation (CV) as the average CV between duplicates for all samples, and we calculated the overall interassay CV across plates using the same pooled test sample (measured at both 'high' and 'low' concentrations, with the specific dilution depending on the analyte measured). We re-ran any samples with a CV that exceeded 15% and only included measurements under that cut-off. Intraassay and interassay CVs, respectively, were 4.5% and 13.2% for the Arbor Assays Cortisol kit; 4.6% and 11.1% for ISWE Cortisol; 6.9% and 16.8% for ISWE Corticosterone; 7.0% and 19.4% for Arbor Assays Testosterone; and 8.2% and 16.2% for ISWE Testosterone. We calculated assay sensitivity by subtracting two standard deviations from the mean optical density value of all zero standard replicates across plates and converting this value into the ng/g concentration reported in our results. These sensitivities are 1.77 ng/g for Arbor Assays Cortisol; 1.60 ng/g for ISWE Cortisol; 8.09 ng/g for ISWE Corticosterone; 3.63 ng/g for Arbor Assays Testosterone; and 1.53 ng/g for ISWE Testosterone.

2.4. Analysis

We evaluated assay performance in several different ways. First, for each of the five 'new' assays, we assessed parallelism and linearity/goodness-of-fit (following e.g. Sheriff et al., 2011) by assaying serial dilutions of a pooled fecal extract. For cortisol assays, the pool contained samples from all age/sex classes, since this assay was previously validated for use with males, females, and a range of ages (Eckardt et al. 2016). For testosterone assays, the pool contained samples only from adult males, since this assay was previously validated for use only with adult males (Rosenbaum et al. 2021). For parallelism, we assessed correspondence of slopes for our serial dilution of mountain gorilla samples compared to serial dilution of kit standards; for linearity and goodness-of-fit, respectively, we report R^2 values for a) linear and b) four-parameter logistic curve fits to a scatter plot of dilution factors and percentage binding values in our EIA results. Our dilution range varied among analytes, based on manufacturer recommendations and previous data in Eckardt et al. (2016) and Rosenbaum et al. (2021) that established the likely range of detectable concentrations for each metabolite. For both cortisol assays, our dilution factor was 2x; as we began with 1:5 samples, to create a suitable assay range, we both serially concentrated these samples to 1:1.25 (by evaporating samples completely and adding assay buffer to the appropriate concentration) and serially diluted them to 1:80. For corticosterone, the dilution factor was also 2x, but the dilution range instead ran from 1:5 to 1:640. For both testosterone assays, we started with 1:5 samples and diluted them by a factor of 2.5x (matching the manufacturer's standards) until our final dilution of 1:1220. Next, to assess inter-assay agreement, we report pairwise Pearson correlations between the measurements obtained from matched samples assayed by each of the FGCM/FAM kits–i.e., we report six pairwise correlation coefficients for FGCMs, and three for FAMs. Finally, to assess ecological validity of our newer assays, we plot concentrations from each of these assays relative to the competitive events assessed in Eckardt et al. (2016) and Rosenbaum et al. (2021), fit these trajectories with a non-linear generalized additive model (GAM; Ross, 2019), and report two figures for each new assay that stem from these models: 1) the significance of the "day relative to event" spline, controlling for animal identity (a "significant" spline in this context can be interpreted as one that cannot have a horizontal line drawn through its 95% confidence interval; see Ross, 2019); and 2) the overall model-estimated proportional increase, calculated by comparing the peak estimated hormone concentration to the average estimated hormone concentration prior to the event.

All data and code necessary to reproduce our results have been posted publicly on the Open Science Framework at https://osf.

io/5eyg6/.

3. Results

3.1. Parallelism and Linearity/Goodness-of-fit

All five assays we tested exhibited acceptable parallelism and linearity/goodness-of-fit (Fig. 1). A comparison of the slope parameters between test sample and standard dilutions revealed no significant differences for any assay (all p greater than 0.05), with the largest divergence in slopes observed for the ISWE Corticosterone kit (Fig. 1C; $t(7) = 2.07$, $p = 0.078$). The Arbor Assays Cortisol kit showed the lowest linear R^2 value between binding and dilution factor of the pooled sample (0.926; Fig. 1A), owing to somewhat poorer binding discrimination at low binding percentages; linear R^2 values for the four other kits all exceeded 0.98 and showed similar resolution across binding percentages. R^2 values for the four-parameter logistic curve fit (blue lines, Fig. 1) exceeded 0.995 for each assay.

3.2. Comparison of newer commercial assays to original FGCM and FAM results

Of the three new FGCM assays tested, the ISWE Cortisol kit correlated most strongly with the original R4866 Cortisol assay ($r = 0.75$; Fig. 2B), followed by the Arbor Assays Cortisol assay ($r = 0.66$; Fig. 2A) and the ISWE Corticosterone assay ($r = 0.53$; Fig. 2D). The new FGCM assays correlated moderately with one another ($r = 0.42 – 0.61$). Correlations were consistently lowest between ISWE Corticosterone measurements and other assays, owing to a handful of especially high corticosterone observations that did not yield similarly high measurements in any cortisol assay (Fig. 2D – 2F). Removing these potential outliers increased ISWE Corticosterone's correlation with other assays (Supplementary Online Materials, Appendix 1), though the rank order of correlations remained consistent.

All three new assays successfully captured the surge in FGCMs following an intergroup encounter originally reported using the R4866 Cortisol assay in Eckardt et al. (2016) (Fig. 3). The non-linear trends depicted in Fig. 3 significantly differed from a flat line for all three new assays: the trend for AA Cortisol corresponded to an estimated 1.67-fold increase ($F = 3.17$, $p = 0.014$; Fig. 3B); the trend for ISWE Cortisol corresponded to an estimated 1.77-fold increase ($F = 6.79$, $p < 0.001$; Fig. 3C); and the trend for ISWE Corticosterone ($F = 8.20$, $p < 0.001$; Fig. 3D) corresponded to a 1.53-fold increase.

Both new assays for FAMs showed very similar, moderately strong correlations with the original R156/7 Testosterone assay ($r = 0.56$ and 0.58); at the same time, these new assays correlated even more strongly with one another ($r = 0.88$). See Fig. 4.

Both new FAM assays captured the increase in androgens following competitive events originally reported using the R156/7 Testosterone assay in Rosenbaum et al. (2021), though there was more uncertainty around the magnitude of this spike compared to FGCMs (Fig. 5). The non-linear trends depicted in Fig. 5 fell just short of significance for both new assays: for AA Testosterone ($F = 2.16$, $p = 0.069$; Fig. 5B) this corresponded to an estimated 1.85-fold increase; for ISWE Testosterone, this corresponded to an estimated 1.75-fold increase ($F = 1.98$, $p = 0.090$; Fig. 3C).

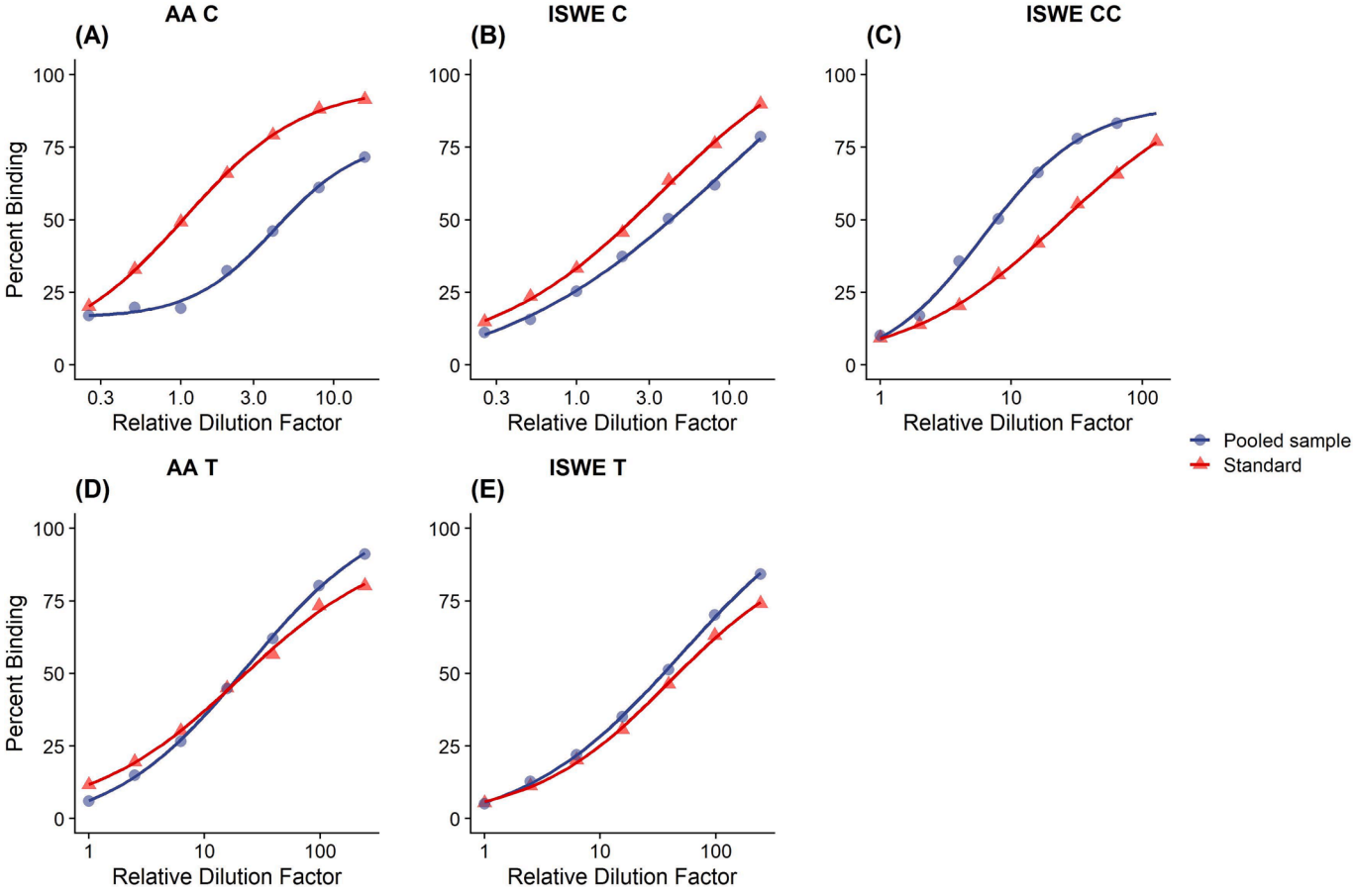


Fig. 1. Parallelism and linearity/goodness-of-fit for each of the five assays evaluated. Parallelism is measured by the similarity of slopes in each plot between blue data points (pooled fecal extracts) and red data points (analytical standards) across a range of dilutions. Linearity/goodness-of-fit is measured by variance explained by a linear trend, or a four-parameter logistic curve, fitted to the blue data points in each plot.

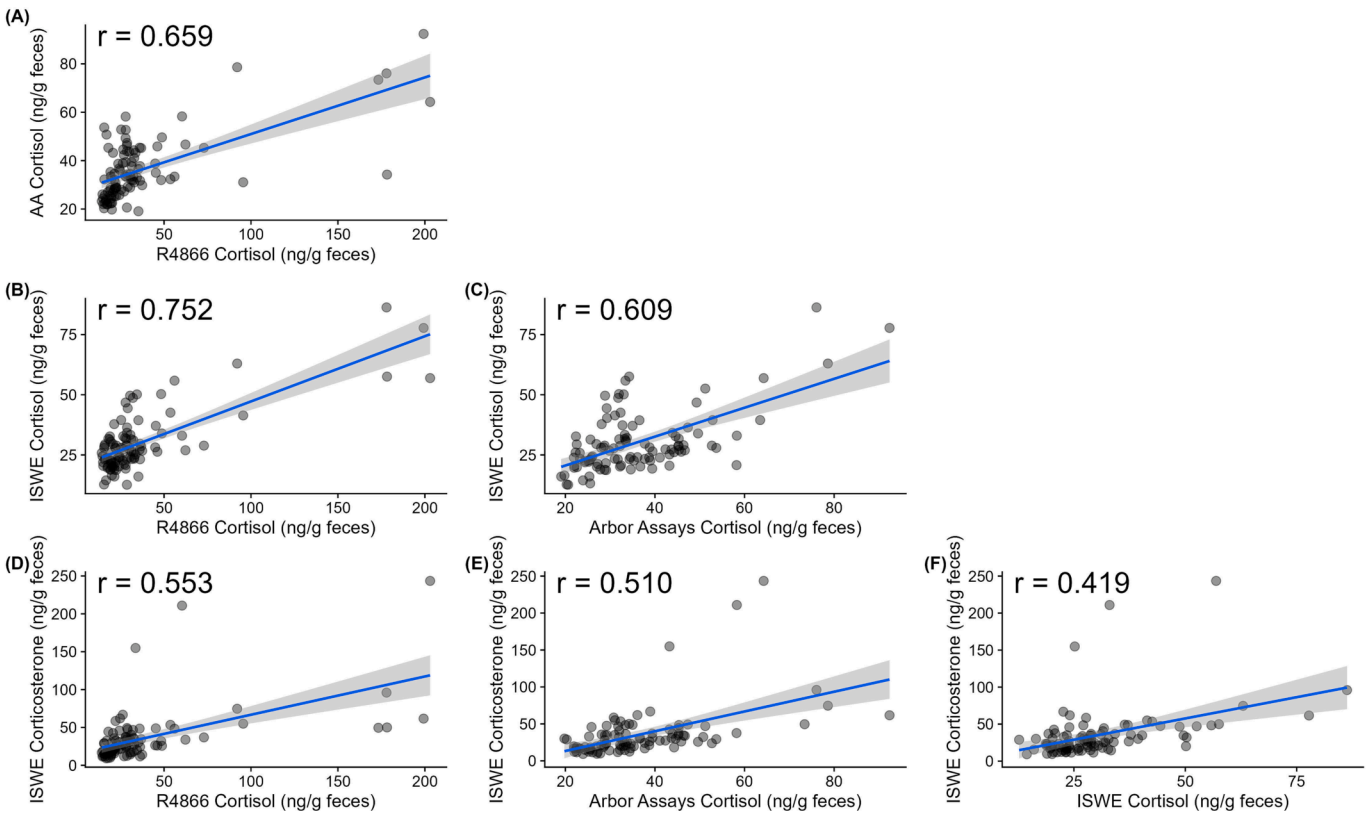


Fig. 2. Pairwise scatter plots depicting and reporting Pearson correlations between matched samples for all available fecal glucocorticoid metabolite immunoassays. Individual data points are plotted at 40% transparency. Shaded regions represent the 95% confidence interval for the line of best fit.

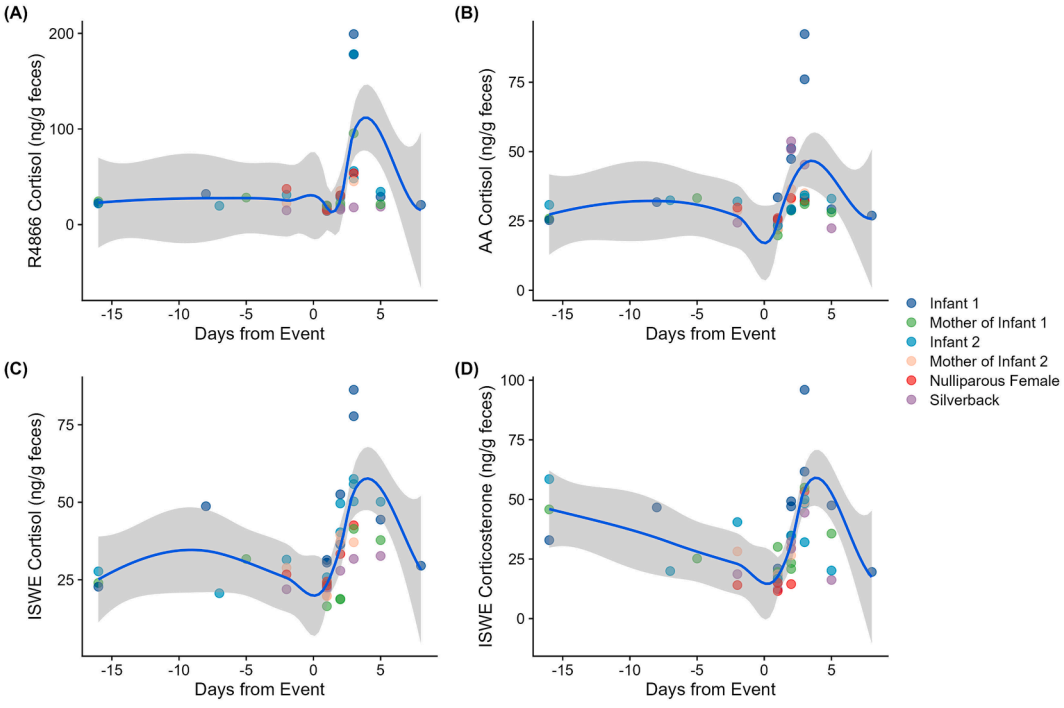


Fig. 3. Non-linear generalized additive models fitted to fecal glucocorticoid metabolite measurements provided by each immunoassay, proximal to individuals experiencing an intergroup encounter. Gray bands indicate 95% confidence intervals for the non-linear spline term for days relative to intergroup encounter.

4. Discussion

Previous guidance on the use of EIAs to measure fecal hormone metabolites has consistently emphasized the need to select assays carefully, as performance can vary substantially between species and populations, and among manufacturers. Thus, the current study was

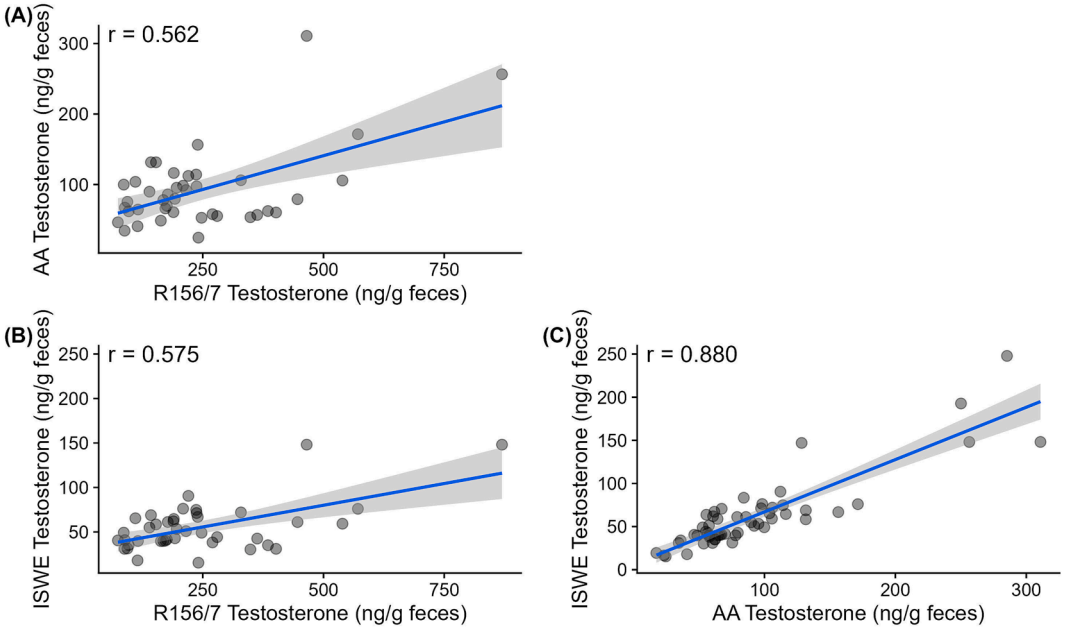


Fig. 4. Pairwise scatter plots depicting and reporting Pearson correlations between matched samples for all available fecal androgen metabolite immunoassays. Individual data points are plotted at 40% transparency. Shaded regions represent the 95% confidence interval for the line of best fit.

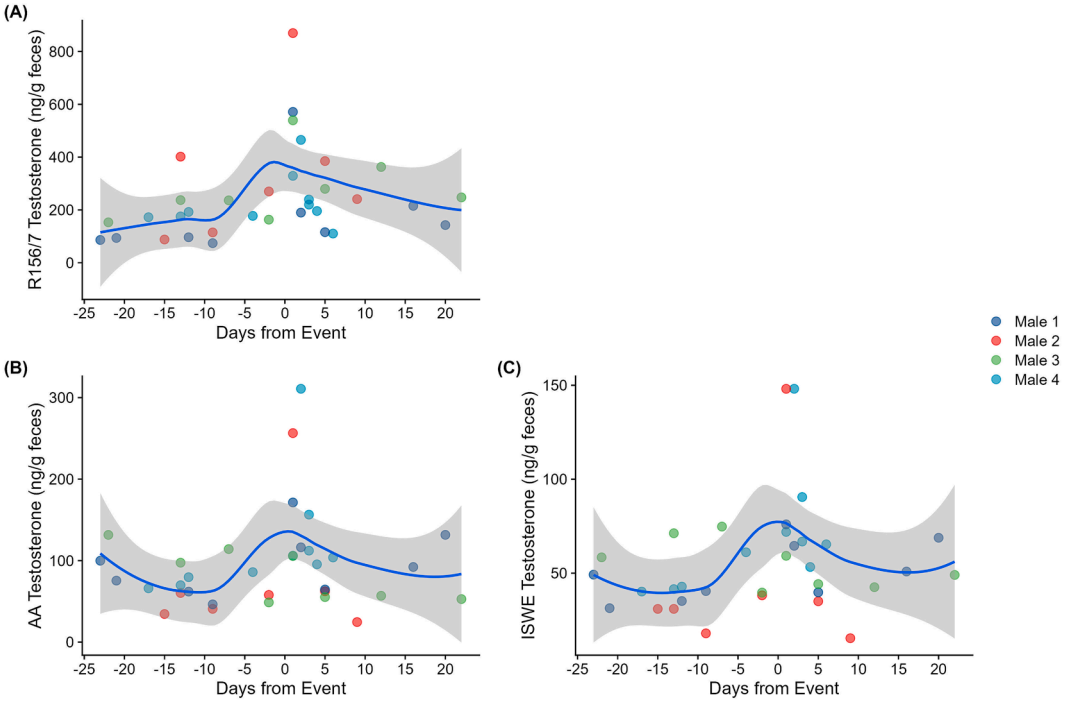


Fig. 5. Non-linear generalized additive models fitted to fecal androgen metabolite measurements provided by each immunoassay, proximal to males experiencing competitive events (including an intragroup fight, an intergroup fight, a female transfer, and leadership of a temporary subgroup; see Rosenbaum et al., 2021). Gray bands indicate 95% confidence intervals for the non-linear spline term for days relative to competitive events.

motivated by two distinct but related needs: more narrowly, to identify appropriate assays for assessing FGCMs and FAMs in wild mountain gorillas; and more generally, to illustrate differences in performance among assays that are each commonly employed in the non-invasive assessment of physiological function.

On one level, our study provides a clear demonstration that, as others have argued (e.g. Touma & Palme, 2005; Hinchcliffe et al., 2021), assay choice affects estimates of gonadal and adrenocortical function obtained from fecal samples. Internal validity checks did not suggest identical parallelism and linearity between kits, even though this divergence fell short of indicating significant deficiencies in the internal validity of any assay. Furthermore, assays correlated imperfectly with one another. At the extreme, consider ISWE Cortisol and ISWE Corticosterone, two assays that have each been used to assess FGCMs in mammals–and in some cases, the same mammalian family (i.e. felids; see ISWE, 2022a and 2022b). These assays only correlated at 0.42 with one another in our analyses, meaning variation in estimates from one kit explained <18% of the variation in the other. While a lack of correspondence may not be

surprising to some researchers, given that these assays are explicitly targeting different parent hormones, it is clear from the published literature that they are sometimes treated interchangeably in practice. For example, research in *Papio* spp. assesses glucocorticoids using both assays for corticosterone (Gesquiere et al., 2008; Habig et al., 2021) and cortisol (Fourie et al., 2015); the same is true for papers on chimpanzees (**corticosterone:** Pizzutto et al., 2015; **cortisol**: Stanton et al., 2015) and orangutans (**corticosterone:** Fink et al., 2022; **cortisol**: Takeshita et al., 2019). Assays marketed to detect the same parent hormone (cortisol or testosterone) generally provided more similar measurements: all matched-sample correlations exceeded 0.5, with an average correlation greater than 0.6. These correlations are comparable to, for example, those reported for matched serum and urinary estradiol measurements in women ($r = 0.54$; Roos et al., 2015). No matter the assay, though, there was still substantial variation in measurements between kits. While we did not test any assays targeted to specific groups of metabolites (e.g., 11-oxoetiocholanolone for FGCMs), we expect that these assays too would vary substantially in estimated concentrations (as shown in e.g. Hinchcliffe et al., 2021).

However, simply demonstrating differences in estimates does not speak to the kinds of issues forefront in the minds of many wildlife researchers. For this audience, ecological validity is likely of greater importance–for instance, does an assay capture physiological changes that are expected to track socioecological variation? Here, one could interpret our results more positively and argue that despite appreciable differences in performance between different FGCM and FAM assays, this noise is fairly marginal compared to the shared signal captured by each assay. While different antibodies possess different sensitivities and specificities, which presents a concern when moving between different immunoassays, the increases in FGCMs and FAMs following competitive events, originally identified by the assays in Eckardt et al. (2016) and Rosenbaum et al. (2021), were also reflected in results from the new assays. Furthermore, we find little evidence that the relative magnitude of these changes varied systematically across kits. In a recent study comparing measurements from two FGCM assays (for corticosterone and 11-oxoetiocholanolone) applied to the same fecal samples in ponies (Hinchcliffe et al., 2021), researchers reported moderately strong positive correlations between assays ($r = 0.56$) when analyzing samples collected across naturally occurring variation in season and social groups. However, in a more restricted comparison of samples collected proximal to the acute stressor of an annual roundup of the population, there was almost no correlation between assays ($r = -0.05$). While our comparisons differ somewhat from Hinchcliffe et al. (2021), as we did not test any assays specifically targeted towards 11-oxoetiocholanolone, our results in mountain gorillas do not suggest a similar breakdown in correspondence: overall correlations between assays were comparable to overall correlations in Hinchcliffe et al. (2021), and we observed similar increases in FGCMs following intergroup encounters in all assays.

Based on our pattern of results, which is consistent with both substantial variation and shared signal, we offer two perspectives for future research: one more narrowly applicable to studies of mountain gorilla physiology, and one that pertains more broadly to wildlife endocrinology.

Recommendations for future studies of fecal hormone metabolites in gorillas. Extant evidence suggests that while all mammals produce detectable amounts of both cortisol and corticosterone, cortisol is the predominant adrenal steroid in primates and corticosterone is generally produced in smaller quantities (Heistermann et al., 2006; Sheriff et al., 2011; Raff, 2016). A comparison of FGCM changes in a lowland gorilla following an ACTH challenge showed that a cortisol EIA outperformed one designed for corticosterone (Heistermann et al., 2006); furthermore, the authors of this paper suggest that an appreciable amount of unmetabolized cortisol, but not corticosterone, was present in gorilla fecal samples. Given this prior evidence and our current findings, we do not recommend the use of corticosterone assays for gorilla fecal samples. Regarding the other FGCM assays, we are unable to confidently address the exact balance of cortisol and its metabolites detected by each. But, evidence of internal and ecological validity for the ISWE Cortisol assay is generally consistent with its advertised ability to measure "cortisol and/or its metabolites" (Arbor Assays, 2021). We conclude this cost-effective assay is an appropriate choice for future studies of gorillas (likely both *G. gorilla* and *G. beringei*, given that Heistermann and colleagues' study was carried out with the former, and ours the latter). The similar performance of the Arbor Assays Cortisol kit in our study suggests it too is an appropriate assay option for these samples. Similarly, the performance of the Arbor Assays and ISWE Testosterone kits was nearly identical in the current study. For future research using gorilla fecal samples, the relatively low per-sample cost of the ISWE kits makes it the more attractive option in many contexts, though the reduced labor associated with Arbor Assays kits may make them an appealing choice for researchers who plan to analyze smaller numbers of samples.

General recommendations for future studies of fecal hormone metabolites. Researchers who are more broadly interested in measuring FGCMs or FAMs, or in synthesizing published studies on these estimates, have a wider set of issues to consider. The most consistent message of methods development research in this domain is that performance can vary substantially across species, manufacturers, and contexts. What works well in one species or situation may not work well in others. Therefore, generally applicable recommendations tend to concern process, rather than specific paths. Many of these recommendations have been voiced previously (e.g., Bahr et al., 2000; Touma & Palme, 2005; Heistermann et al., 2006; Palme, 2019; Hinchcliffe et al., 2021), but as our results reinforce, they are evergreen.

Authors of *meta*-analyses or reviews will need to contend with the fact that differences in hormone metabolite concentration estimates from individual studies, even on the same study species and/or sample type, may be in part attributable due to interassay variation rather than true physiological differences. Concentrations, therefore, must always be standardized (e.g., by log-transforming and *z*-scoring) within assays when making any quantitative comparisons across assays. This point is also relevant to any laboratories, like ours, that need to switch assay manufacturers during the process of longitudinal hormonal monitoring of individuals. Researchers seeking to establish or modify a program of non-invasive monitoring in their study species should view the validity of fecal hormone metabolite EIAs as a key element of study design. Ideally, validations of an assay in the same study species, or a closely related one, will have already been performed, which is more likely than ever thanks to the efforts of wildlife endocrinologists. Having this evidence base provides an invaluable starting point, but even in such an ideal situation, internal validity checks are necessary to ensure the assay performs as advertised in new samples. Absent pre-existing validation of assays, researchers will need to generate them by examining inter-assay agreement and performance against physiological or socioecological benchmarks. We have provided an example of one such validation with our present study. These efforts will be perennially necessary to ensure the biological relevance of EIA measurements for hormone metabolites across a diverse range of species, and to promote the continued growth of wildlife endocrinology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are publicly available in an OSF repository (https://osf.io/5eyg6/)

Acknowledgments

We thank the Rwandan government and the Rwanda Development Board for their long-term support of the research, monitoring, and protection activities of the Dian Fossey Gorilla Fund. We are deeply indebted to all Fossey Fund field staff for their tireless support in collecting long-term behavioral and demographic data. This study was supported by the University of Michigan, the NSF Graduate Research Fellowship Program and Doctoral Dissertation Improvement Grant No. 1122321, and the donors who support the Dian Fossey Gorilla Fund.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygcen.2023.114351.

References

Arslan, R.C., Blake, K., Botzet, L.J., Bürkner, P.-C., DeBruine, L., Fiers, T., Grebe, N., Hahn, A., Jones, B.C., Marcinkowska, U.M., Mumford, S.L., Penke, L., Roney, J.R., Schisterman, E.F., Stern, J., 2023. Not within spitting distance: Salivary immunoassays of estradiol have subpar validity for predicting cycle phase. Psychoneuroendocrinology 149, 105994.

Bahr, N.I., Palme, R., Möhle, U., Hodges, J.K., Heistermann, M., 2000. Comparative aspects of the metabolism and excretion of cortisol in three individual nonhuman primates. General Comp. Endocrinol. 117 (3), 427–438.

Beehner, J.C., Whitten, P.L., 2004. Modifications of a field method for fecal steroid analysis in baboons. Physiol. Behav. 82 (2–3), 269–277.

Behringer, V., Deschner, T., 2017. Non-invasive monitoring of physiological markers in primates. Hormones Behav. 91, 3–18.

Eckardt, W., Stoinski, T.S., Rosenbaum, S., Umuhoza, M.R., Santymire, R., 2016. Validating faecal glucocorticoid metabolite analysis in the Virunga mountain gorilla using a natural biological stressor. Conserv. Physiol. 4 (1), cow029.

Eckardt, W., Stoinski, T.S., Rosenbaum, S., Santymire, R., 2019. Social and ecological factors alter stress physiology of Virunga mountain gorillas (*Gorilla beringei beringei*). Ecol. Evol. 9 (9), 5248–5259.

Fink, L.B., Mukobi, A., Gruber, L., Reed, C., DeLibero, J., Jackson, S., Neill, S., Walz, J., Sines, C., VanBeek, B., Scarlata, C.D., Wielebnowski, N., 2022. Longitudinal analysis of variability in fecal glucocorticoid metabolite concentrations in three orangutans (*Pongo pygmaeus pygmaeus and Pongo pygmaeus abelii*) before, during, and after transition from a regular habitat environment to temporary housing in indoor holding facilities. Animals 12 (23), 3303.

Fourie, N.H., Jolly, C.J., Phillips-Conroy, J.E., Brown, J.L., Bernstein, R.M., 2015. Variation of hair cortisol concentrations among wild populations of two baboon species (*Papnubisbis, P. hamadryas*) and a population of their natural hybrids. Primates 56, 259–272.

Freeman, E.W., Meyer, J.M., Bird, J., Adendorff, J., Schulte, B.A., Santymire, R.M., 2014. Impacts of environmental pressures on the reproductive physiology of subpopulations of black rhinoceros (*Diceros bicornis bicornis*) in Addo Elephant National Park, South Africa. Conserv. Physiol. 2 (1), cot034–cot.

Ganswindt, A., Brown, J.L., Freeman, E.W., Kouba, A.J., Penfold, L.M., Santymire, R.M., Vick, M.M., Wielebnowski, N., Willis, E.L., Milnes, M.R., 2012. International Society for Wildlife Endocrinology: the future of endocrine measures for reproductive science, animal welfare and conservation biology. Biol. Lett. 8 (5), 695–697.

Gelman, A., Loken, E., 2014. The statistical crisis in science data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. Am. Sci. 102 (6), 460.

Gesquiere, L.R., Khan, M., Shek, L., Wango, T.L., Wango, E.O., Alberts, S.C., Altmann, J., 2008. Coping with a challenging environment: effects of seasonal variability and reproductive status on glucocorticoid concentrations of female baboons (Papio cynocephalus). Horm. Behav. 54 (3), 410–416.

Habig, B., Chowdhury, S., Monfort, S.L., Brown, J.L., Swedell, L., Foerster, S., 2021. Predictors of helminth parasite infection in female chacma baboons (Papio ursinus). Int. J. Parasitol: Para. Wildlife 14, 308–320.

Heistermann, M., Palme, R., Ganswindt, A., 2006. Comparison of different enzymeimmunoassays for assessment of adrenocortical activity in primates based on fecal analysis. Am. J. Primatol. 68 (3), 257–273.

Hinchcliffe, D.L., Lea, J.M., Palme, R., Shultz, S., 2021. Fecal glucocorticoid metabolites as biomarkers in equids: assay choice matters. J. Wildl. Manag. 85 (6), 1175–1186.

ISWE Corticosterone Mini-Kit. Arbor Assays. (2022, March 22). Retrieved July 1, 2022, from https://www.arborassays.com/documentation/inserts/ISWE007.pdf.

ISWE Cortisol Mini-Kit. Arbor Assays. (2021, July 7). Retrieved July 1, 2022, from https://www.arborassays.com/documentation/inserts/ISWE002.pdf.

King, A.G., Edwards, P.D., Cote, S., Palme, R., Boonstra, R., Sicotte, P., 2023. Assessing stress in wild black-and-white colobus monkeys non-invasively. Gen. Comp. Endocrinol. 334, 114212.

Loeding, E., Thomas, J., Bernier, D., Santymire, R., 2011. Using fecal hormonal and behavioral analyses to evaluate the introduction of two sable antelope at Lincoln Park Zoo. J. Appl. Anim. Welf. Sci. 14 (3), 220–246.

Murray, C.M., Heintz, M.R., Lonsdorf, E.V., Parr, L.A., Santymire, R.M., 2013. Validation of a field technique and characterization of fecal glucocorticoid metabolite analysis in wild Chimpanzees (*Pan troglodytes*). Am. J. Primatol. 75 (1), 57–64.

Murtagh, R., Behringer, V., Deschner, T., 2013. LC-MS as a method for non-invasive measurement of steroid hormones and their metabolites in urine and faeces of animals. Wien Tierärztl Monat-Vet Med Austria 100, 247–254.

Palme, R., 2005. Measuring fecal steroids: guidelines for practical application. Ann. N. Y. Acad. Sci. 1046 (1), 75–80.

Palme, R., 2019. Non-invasive measurement of glucocorticoids: Advances and problems. Physiol. Behav. 199, 229–243.

Palme, R., Touma, C., Arias, N., Dominchin, M.F., Lepschy, M., 2013. Steroid extraction: get the best out of faecal samples. Wien. Tierarztl. Monatsschr. 100 (9/10), 238–246.

Pappano, D.J., Roberts, E.K., Beehner, J.C., 2010. Testing extraction and storage parameters for a fecal hormone method. Am. J. Primatol. 72 (11), 934–941.

Pizzutto, C.S., Sgai, M.G.F.G., Lopes, D.A., Pessutti, C., Nunes, A., Furtado, P.V., Oliveira, C.A.d., Guimarães, M.A.B.V., 2015. Relation between the level of self-mutilation and the concentration of fecal metabolites of glucocorticoids in captive chimpanzees (Pan troglodytes). Pesquisa Veterinária Brasileira 35 (1), 62–66.

Pollastri, I., Normando, S., Florio, D., Ferrante, L., Bandoli, F., Macchi, E., Muzzo, A., de Mori, B., Yue, B.-S., 2022. The animal-visitor interaction protocol (AVIP) for the assessment of Lemur catta walk-in enclosure in zoos. PLoS One 17 (7), e0271409.

Raff, H., 2016. CORT, Cort, B, corticosterone, and now cortistatin: enough already! Endocrinology 157 (9), 3307–3308.

Romano, M.C., Rodas, A.Z., Valdez, R.A., Hernández, S.E., Galindo, F., Canales, D., Brousset, D.M., 2010. Stress in wildlife species: noninvasive monitoring of glucocorticoids. Neuroimmunomodulation 17 (3), 209–212.

Roos, J., Johnson, S., Weddell, S., Godehardt, E., Schiffner, J., Freundl, G., Gnoth, C., 2015. Monitoring the menstrual cycle: Comparison of urinary and serum reproductive hormones referenced to true ovulation. Eur. J. Contracept. Reprod. Health Care 20 (6), 438–450.

Rosenbaum, S., Eckardt, W., Stoinski, T.S., Umuhoza, R., Kuzawa, C.W., Santymire, R.M., 2021. Group structure, but not dominance rank, predicts fecal androgen metabolite concentrations of wild male mountain gorillas (Gorilla beringei beringei). Am. J. Primatol. 83 (8), e23295.

Ross, N. (2019, May 30). Generalized additive models in r · a free interactive course. Retrieved September 14, 2022, from https://noamross.github.io/gams-in-r-course/.

Santymire, R.M., Armstrong, D.M., 2010. Development of a field-friendly technique for fecal steroid extraction and storage using the African wild dog (Lycaon pictus). Zoo Biology 29 (3), 289–302.

Schwarzenberger, F., 2007. The many uses of non-invasive faecal steroid monitoring in zoo and wildlife species. Internat. Zoo Yearbook 41 (1), 52–74.

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.

Stanton, M.A., Heintz, M.R., Lonsdorf, E.V., Santymire, R.M., Lipende, I., Murray, C.M., 2015. Maternal behavior and physiological stress levels in wild chimpanzees (Pan troglodytes schweinfurthii). Int. J. Primatol. 36 (3), 473–488.

Takeshita, R.S.C., Mendonça, R.S., Bercovitch, F.B., Huffman, M.A., 2019. Developmental changes in the endocrine stress response in orangutans (Pongo pygmaeus). J. Comp. Physiol. B 189 (6), 659–672.

Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N. Y. Acad. Sci. 1046 (1), 54–74.

Welker, K.M., Lassetter, B., Brandes, C.M., Prasad, S., Koop, D.R., Mehta, P.H., 2016. A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry. Psychoneuroendocrinology 71, 180–188.

Wielebnowski, N., Watters, J., 2007. Applying fecal endocrine monitoring to conservation and behavior studies of wild mammals: important considerations and preliminary tests. Israel J. Ecol. Evolut. 53 (3–4), 439–460.