Questioning the utility of oxidative stress measurements as biomarkers of physiological condition and fitness

Rachel R Reid[1*], Davide M Dominoni[1 *], Jelle Boonekamp[1]

[1]School of Biodiversity, One Health and Veterinary Medicine, Graham Kerr Building, University of Glasgow, Glasgow, G12 8QQ, UK

[*]Email: Rachelrosereid141@outlook.com

Keywords: Oxidative stress, intraindividual variation, measurement reliability, ageing physiology

Abstract

Markers of oxidative stress are widely used as biomarkers of health, ageing, and physiological stress. However, their reliability as biomarkers remains uncertain due to high intraindividual variation, obscuring associations with environmental conditions, lifestyle, frailty, and physiological indicators of health such as telomere length. Whilst numerous longitudinal studies exist, the individual repeatability of oxidative stress measurements is rarely reported. This study presents the first meta-analysis assessing individual repeatability of oxidative stress markers, comprising 123 repeatability estimates obtained from 22 studies. We found that oxidative stress exhibits low individual repeatability (Intraclass correlation = 0.164), regardless of oxidative stress marker type, taxa, sex, study design, or environment. This flags serious limitations regarding the utility of oxidative stress measurements to meaningfully capture aspects of health and to predict other health-related traits such as telomere length. In line with this assertion, we found that different markers of oxidative stress were often poorly correlated. Next, we simulated causal effects of oxidative stress on telomere length to reveal statistical power limitations on the detection of a relationship between oxidative stress and telomere length when the individual repeatability of oxidative stress is consistent with our meta-analysis findings. Our simulations reveal that substantially larger samples sizes are required than those typically used in this field. On a more positive note, we also show that increasing the number of repeated samples can improve statistical power. This would create valuable opportunities for untangling the causes of intraindividual variation in oxidative stress, improving the utility of oxidative stress as biomarkers of environmental effects on fitness, as well as elucidating the functional consequences of oxidative stress for health and ageing.

Introduction

Oxidative stress is a pervasive feature of life, implicated in cellular senescence, ageing, and disease (1). It arises when prooxidants, by-products of aerobic metabolism, exceed the neutralising capacity of antioxidants (2–4). This imbalance initiates a cascade of reactions causing DNA damage, alterations to membrane lipids, and damage to other macro-molecules (5). On a systemic level such cellular damage is widely believed to contribute to organismal senescence (6–8) and ageing related disease (9–12). Understanding the sources of individual variation in oxidative stress may therefore offer crucial insights into the mechanisms of ageing and health. For example, elevated levels of oxidative stress might reveal an unhealthy lifestyle if oxidative stress measurements link lifestyle factors to health. However, health status beyond obvious disease is difficult to measure. Therefore, the utility of oxidative stress as biomarkers depends on their associations with other health markers such as telomere length. Herein lies a fundamental challenge as oxidative stress is often poorly correlated with other biomarkers. For example, although telomere length is widely believed to be linked with adult mortality (13) and oxidative stress (5), a recent meta-analysis (14) found that the correlation between oxidative stress and telomere length is surprisingly weak (r = 0.027). This raises important concerns about whether markers of oxidative stress capture meaningful aspects of health (14). This issue is of particular importance given the widespread use of oxidative stress biomarkers in clinical and pharmaceutical research (15, 16), ecotoxicology (17) and evolutionary ecology (2, 18, 19). The cause of the weak association between oxidative stress and telomere length remains unclear. One simple explanation could be that oxidative stress markers show high intraindividual variability such that the observed level of oxidative stress at a given point in time might poorly reflect the oxidative stress profile of that individual. This is especially problematic when the intraindividual variation is unpredictable. Hence, the *conditional* repeatability (controlling statistically for potential sources of variation in the data (20)) fundamentally underpins the utility of oxidative stress measurements as biomarkers of health and fitness (Fig. 1).

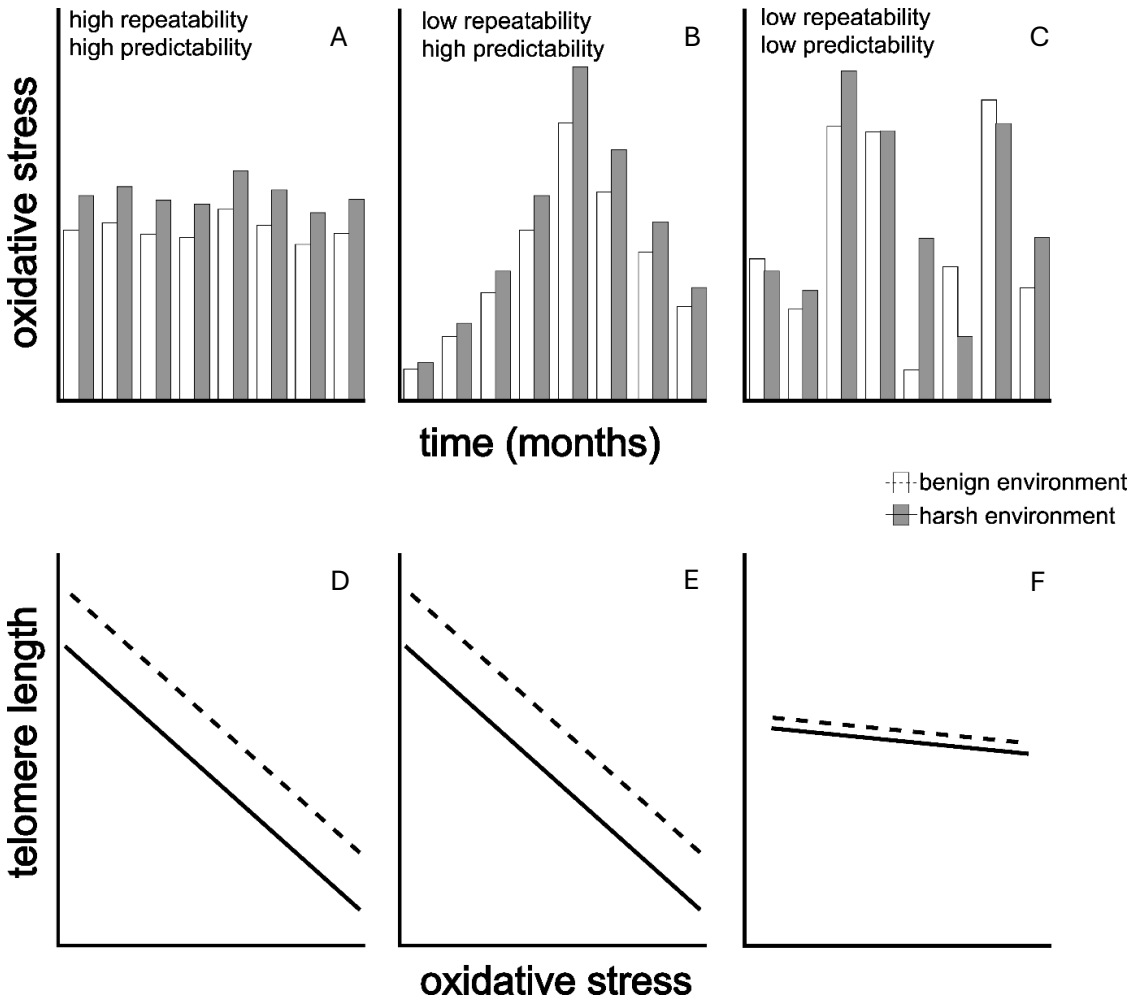

Figure 1. Conceptual depiction of the impact of different scenarios of intraindividual variation in oxidative stress on their utility to predict telomere length in benign and harsh environments. Panels A-C depict repeated individual measurements of oxidative stress across time for (A) high repeatability and predictability, (B) low repeatability and high predictability, (C) low repeatability and low predictability for an individual living in a benign (white bars), and another individual living a harsh environment (grey bars). Panels D-F illustrate the impact of these corresponding scenarios on the population-level associations between oxidative stress, telomere length, and environmental condition. Panel E illustrates that there is little negative impact of low repeatability when one can statistically control for the confounding variation, in this case month. However, oxidative stress is a poor biomarker in the third scenario (panels C and F) because it fails to meaningfully link the environmental effect on oxidative stress to telomere length. Following the logic set out in these simplistic conceptual scenarios, the utility of oxidative stress measurements as health biomarkers hinges on their *conditional* repeatability, i.e., the repeatability *after* successfully controlling for confounding variables.

While establishing individual repeatability has been a key research focus for several physiological traits including metabolic rate (21), glucocorticoid hormones (22, 23), cognitive performance (24), and telomere length (25), the individual repeatability of oxidative stress markers remains unclear. Where studies do report them, repeatability estimates evidently vary widely. For example, glutathione peroxidase displayed moderate repeatability (r = 0.47) in great tits (*Parus major*) across seasons (26). In collared flycatchers' (*Ficedula albicollis*) reactive oxygen metabolites were not repeatable (r = 0.03), whereas antioxidant defences exhibited moderate repeatability (r = 0.581) (27). A growing body of evidence suggests that the individual repeatability of oxidative stress can be affected by a range of environmental and life-history factors. For example, changes in oxidative stress levels have been observed across life-history stages (2, 28, 29) and has been shown to be elevated during mating and reproduction (30, 31), and high-energy activities like migration (32). Oxidative stress is also affected by certain environmental conditions including predation risk (33), territory quality (34), and seasonal changes (35, 36). The plethora of factors that can influence oxidative stress raises the challenge of

how to control for this multitude of environmental effects on intraindividual variation in oxidative stress. Further adding to this challenge, oxidative stress has been shown to fluctuate over very short timescales (37, 38), raising concerns about their fine scale temporal stability. From a 'biomarker utility' perspective, the question remains, to what extent is the observed intraindividual variation predictable? Are there universal factors that underpin intraindividual variation in oxidative stress? Are there certain types of oxidative stress markers that might be more stable and therefore more suitable as a biomarker? Certain antioxidants like vitamin E will largely reflect diet (39, 40) whereas more 'pure' measurements of reactive oxygen species might reflect more dynamic physiological processes linked to metabolism. There could also be differences in the transient behaviour of enzymatic versus non-enzymatic, and intra- or extracellular markers of oxidative stress. Indeed, different oxidative stress markers are often very poorly correlated, even within the same individual (3, 12, 41–43). These issues highlight a need to quantify intraindividual variation across a panel of commonly used oxidative stress markers, to elucidate the potential causes of this variation, and to determine whether certain markers of oxidative stress might be more suitable as biomarkers than others.

The increasing availability of longitudinal studies offers the opportunity to investigate the level of individual repeatability of oxidative stress markers for a wide range of species and identify factors contributing to intraindividual variation. Here, we present the first formal meta-analysis of individual repeatability of oxidative stress. By examining biological and methodological factors, we aim to estimate the overall repeatability and to identify the causes of variation in repeatability. We also evaluate the correlations among oxidative stress markers to determine if any systematic/conserved patterns exist. We subsequently explore the consequences of the observed low repeatability for study design and statistical power using individual based simulations on the association between oxidative stress and telomere length and provide recommendations addressing these issues for future studies.

Results

Our meta-analysis synthesises 123 estimates of individual repeatability, estimated with the intraclass correlation (ICC), derived from 22 published datasets. Across all studies, the overall repeatability of oxidative stress markers was low (ICC = 0.164, 95% confidence interval (CI) = [0.093, 0.233], $P < 0.001$; Table S1, Fig.2), with substantial heterogeneity among effect sizes ($I^2 = 57.4$). Mammals exhibited slightly higher ICC values (ICC = 0.212, 95% CI = [0.058, 0.356], P = 0.007) compared to birds (ICC = 0.155, 95% CI = [0.062, 0.245], P = 0.012; Table S1, Fig.2), but the difference between groups was small and not statistically significant (estimate = 0.059, 95% CI = [-0.123, 0.238], P = 0.526). These findings indicate that oxidative stress markers show low repeatability overall, regardless of taxonomic group. The high heterogeneity observed among effect sizes is noteworthy and indicative of a range of repeatability estimates including instances of moderate to high repeatability. This raises the question whether there could be common factors underpinning the higher repeatability observed in a minority of cases, and vice-versa whether there are systematic explanations for low repeatability observed in the majority of studies.

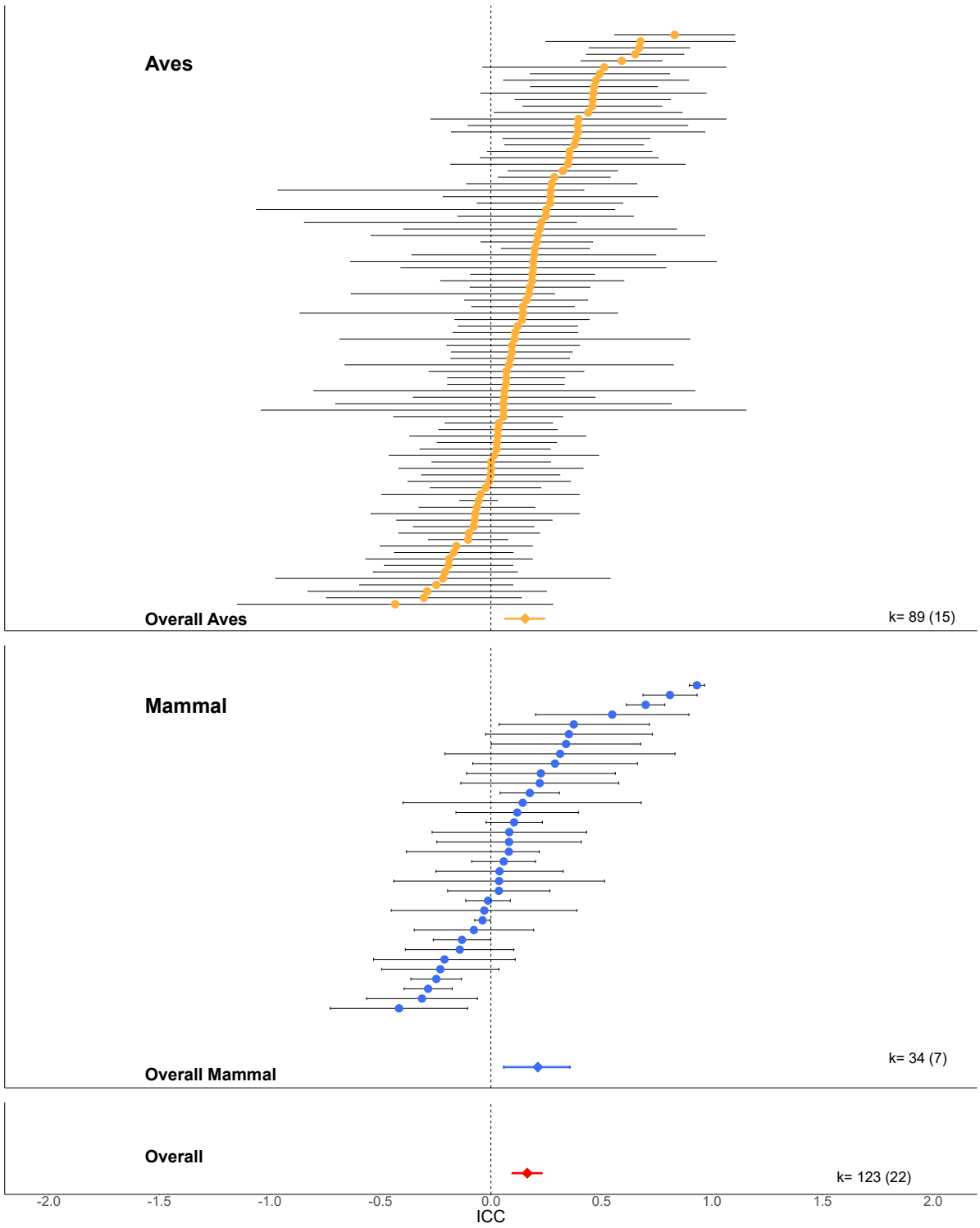

Figure 2: Caterpillar plot displaying individual effect sizes sorted from lowest to highest repeatability and split by taxonomic class. *Yellow points represent ICC values for the "Aves" class while blue points correspond to ICC values for the "Mammal" class. The global ICC estimate for "Aves" is shown at the bottom of the Aves class facet, and the global estimate for "Mammals" is shown similary. These estimates were extracted from the meta-model where taxonomic class was included as a moderator. ICC values are presented along with their 95% confidence intervals. The dashed line represents an ICC value of zero. The red diamond and error bars at the bottom of the plot represents the overall ICC value and its 95% confidence intervals from the intercept only model. The corresponding number of effect sizes is shown on the right side of the plot, with the number of studies indicated in parentheses.*

Moderator analyses investigating the potential causes of low repeatability

We estimated ICC within treatment and control groups for experimental studies, and within sex to enable testing the influence of these variables as moderators in our meta-analysis (see methods for details). The control groups of experimental studies were pooled with the effect sizes of correlational studies to create a 2-level moderator comprising manipulated animals and unmanipulated animals. Contrary to our prediction, effect sizes obtained from unmanipulated animals showed slightly lower repeatability (ICC = 0.146, 95% CI = [0.053, 0.230]) than effect sizes from animals exposed to experimental manipulations (ICC = 0.208, 95% CI = [0.091, 0.319]), but the difference was not significant (estimate = -0.064, 95% CI = [-0.204, 0.077], P = 0.86; Table S1, Fig. 3A). Females also showed similar repeatability (ICC = 0.167, 95% CI = [0.065, 0.266]) to males (ICC = 0.154, 95% CI = [0.03, 0.273]; estimate = -0.013, 95% CI = [-0.163, 0.137], P = 0.86; Fig. 3A-B; Table S1, Fig. 3B).

We observed higher repeatability for studies on livestock populations (ICC = 0.391, 95% CI = [0.178, 0.569], P = 0.005), compared to studies using wild (ICC = 0.119, 95% CI = [-0.003, 0.239], P = 0.057), wild captive (ICC = 0.165, 95% CI = [-0.052, 0.367], P = 0.136), or captive-bred animals (ICC = 0.158, 95% CI = [-0.07, 0.371], P = 0.173; Table S1, Fig. 3C). Using livestock as the reference category, model estimates indicated significantly lower repeatability in wild populations (estimate = -0.285, 95% CI = [-0.506, -0.029], P = 0.029). Repeatability was lower in wild-captive (estimate = -0.242, 95% CI = [-0.513, 0.072], P = 0.129), and captive populations (estimate = -0.248, 95% CI = [-0.523, 0.074], P = 0.074) but neither were significantly different from the livestock reference level. The elevated repeatability in livestock must be interpreted with caution, as these datasets included a higher number of repeated measures per individual, which can increase ICC estimates.

We found no evidence for systematic differences in repeatability among markers of oxidative damage versus antioxidant protection (estimate = -0.016, 95% CI = [-0.140, 0.109], P = 0.803). Both categories showed low but significant repeatability: oxidative damage (ICC = 0.156, 95% CI = [0.056, 0.253], P = 0.002); antioxidant defence (ICC = 0.171, 95% CI = [0.082, 0.258], P = <0.001; Table S1, Fig.3D). Further analysis of oxidative damage marker subtypes revealed that only lipid peroxidation markers exhibited moderate repeatability (ICC = 0.368, 95% CI = [0.182, 0.529], P <0.001; Table S2, Fig.3E). Other damage types, including DNA damage (ICC = 0.135, 95% CI = [-0.224, 0.461], P = 0.464), protein damage (ICC = -0.068, 95% CI = [-0.346, 0.220], P = 0.646), and nonspecific damage (ICC = 0.059, 95% CI = [-0.137, 0.252], P = 0.59), were not statistically significant. Using lipid peroxidation as a reference level, protein damage (estimate = -0.426, 95% CI = [-0.654, -0.128], P = 0.006) and non-specific damage (estimate = -0.316, 95% CI = [-0.539, -0.051], P = 0.020) were shown to have significantly lower repeatability. DNA damage also had lower repeatability than lipid peroxidation, but this difference was not significant (estimate = -0.245, 95% CI = [-0.578, 0.156], P = 0.228). In a similar way, enzymatic antioxidants showed slightly higher repeatability (ICC = 0.204, 95% CI = [0.081, 0.320], P = 0.001), compared to non-enzymatic defence markers (ICC = 0.136 95% CI = [0.042, 0.228], P = 0.005; Table S2, Fig.3F), but the difference between these two categories was not statistically significant (estimate = -0.069, 95% CI = [-0.198, 0.061], P = 0.296), and both subtypes showed low repeatability overall.

Finally, we tested whether the mean time elapsed between repeated oxidative stress measurements and the number of repeated samples per individual influenced estimates of within-individual repeatability, as we would expect these to reduce and increase repeatability, respectively. Surprisingly, the average time interval between repeated measurements had a small but statistically significant positive effect on ICC (slope estimate = 0.022, 95% CI = [0.006, 0.038], P = 0.008; Table S1, Fig. 3G). In line with expectations, the number of repeated measurements per individual had a significant positive effect on ICC (slope estimate = 0.097, 95% CI = [0.064, 0.130], P <0.001; Table S1, Fig. 3H).

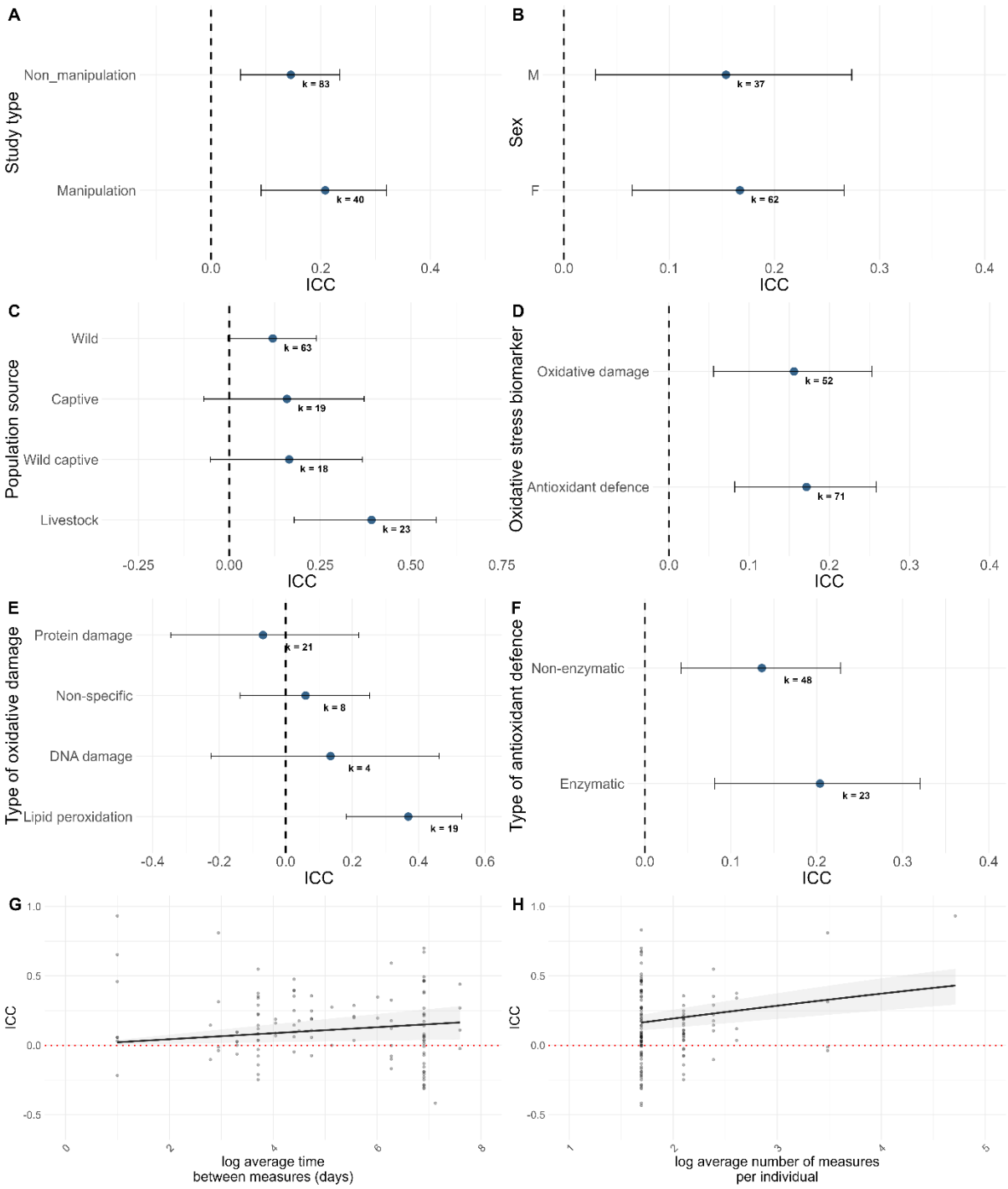

Figure 3: Within-individual repeatability of oxidative stress measures for each moderator level. *Panels A-F show repeatability estimates for categorical moderators (A) Study type, (B) Sex, (C) Population source, (D) Oxidative stress biomarker, (E) Type of oxidative damage, (F) Type of antioxidant defence. For each level, the model-estimated repeatability (ICC) is plotted with 95% confidence intervals, back transformed from Fishers Z. The dashed vertical line indicates an ICC of zero. Panels G and H display the relationship between continuous moderators and repeatability: (G) Log-transformed average time between measures and (H) Log-transformed average number of measures per individual. In each panel, grey points represent individual effect sizes, the black line shows the model-predicted relationship, and the shaded ribbon denotes the 95 confidence intervals. The red dashed line indicates an ICC = 0. ICC is plotted on the y-axis in G and H, and on the x axis in A-F.*

Conditional repeatability analysis

To assess whether controlling for individual-level covariates enhanced repeatability estimates of oxidative stress biomarkers, we conducted a separate meta-analysis using raw data from eight studies that included information on age, to compute conditional repeatability on age. We explored if a similar approach would have been possible for other variables, but unfortunately there was limited information available from the 22 studies at our disposal. For each study we calculated repeatability estimates using four model types: (1) raw (no covariates), (2) age adjusted, (3) age and sex adjusted (where data on sex was available), (4) age and treatment adjusted (where data on treatment was available). Across all models, estimates of repeatability were broadly similar (Fig. 4). Raw repeatability (R = 0.195, 95% CI = [0.067, 0.316], P = 0.003), and conditional repeatability estimates controlling for age (R = 0.187, 95% CI = [0.058, 0.309], P = 0.004), or age and sex (R = 0.167, 95% CI = [0.025, 0.302], P = 0.021) were all indistinguishable from each other. However, conditional repeatability adjusting for age and treatment (R = 0.356, 95% CI = [-0.035, 0.653], P = 0.073, Fig.4) yielded a higher repeatability estimate although this was only based on two effect sizes. These findings suggest that controlling for common covariates such as age and sex has minimal impact on repeatability estimates. This reinforces the conclusion that oxidative stress markers tend to show low repeatability across contexts and that we still have little understanding of the drivers' causing variation in oxidative stress.

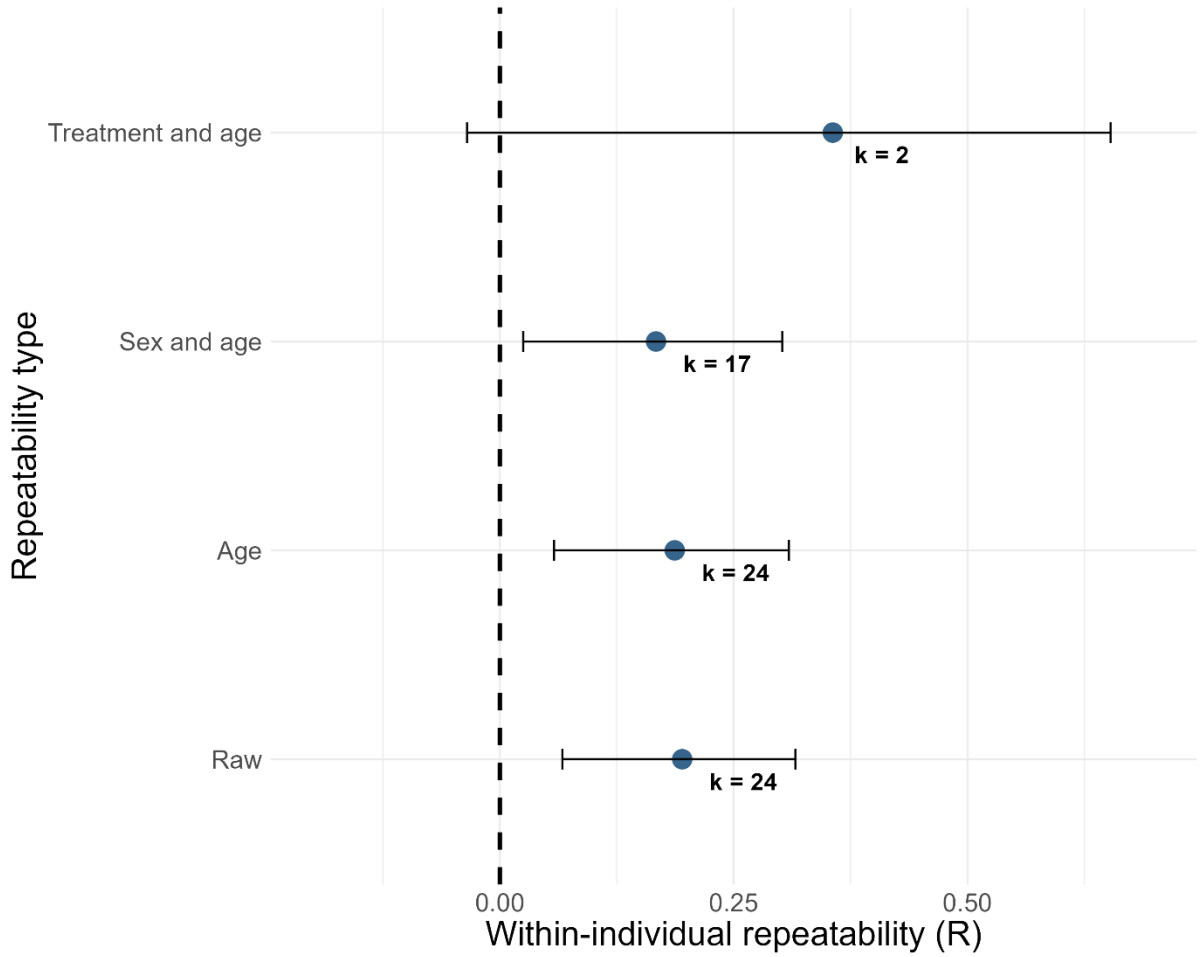

Figure 4: Within-individual repeatability of oxidative stress measures for the type of repeatability calculation. *Repeatability type refers to what variables were controlled for in the model when estimating repeatability. Model estimates are shown along with the 95% confidence intervals. The dashed line indicates an R value of zero. The moderator is shown on the y axis, and R values (back transformed from Fishers Z values) are plotted on the x-axis.*

Are biomarkers of oxidative stress correlated?

Oxidative stress biomarkers are often assumed to reflect interrelated physiological processes, yet studies frequently report weak correlations among them (e.g., 24, 37). Our meta-analysis, showing generally low within-individual repeatability for oxidative stress markers further challenges this view. From a statistical point of view, given an average repeatability of 20%, the maximum average correlation between two markers would be r = 0.04, underscoring the multiplicative impact low repeatability has on the expected correlations among different markers of oxidative stress. Using individual-level raw data available from included studies, we were able to directly assess the correlations between commonly measured oxidative stress markers, including the following pairs of markers; protein carbonyls – superoxide dismutase (carbonyl-SOD), malondialdehyde – antioxidant capacity (MDA-OXY), antioxidant capacity – superoxide dismutase (OXY-SOD), malondialdehyde – superoxide dismutase (MDA-SOD), and reactive oxygen metabolites – antioxidant capacity (dROMs-OXY) (Fig. 5). In accordance with their low repeatability, most pairwise correlations were weak and statistically non-significant (Table S3). However,, we found a significant and modest positive correlation between dROMs and OXY (r = 0.250, 95% CI = [0.091, 0.396]; Table S3, Fig.4). This finding is noteworthy because the low ICC values of dROMs and OXY (0.107 and 0.133, respectively) would predict a much weaker correlation (expected r = 0.014). This could suggest that both markers show synchronous temporal variation within individuals. Nevertheless, the overall pattern reinforces the idea that oxidative stress markers are poorly related and that longitudinal and multivariate approaches are required to capture meaningful aspects of *in vivo* oxidative physiology.

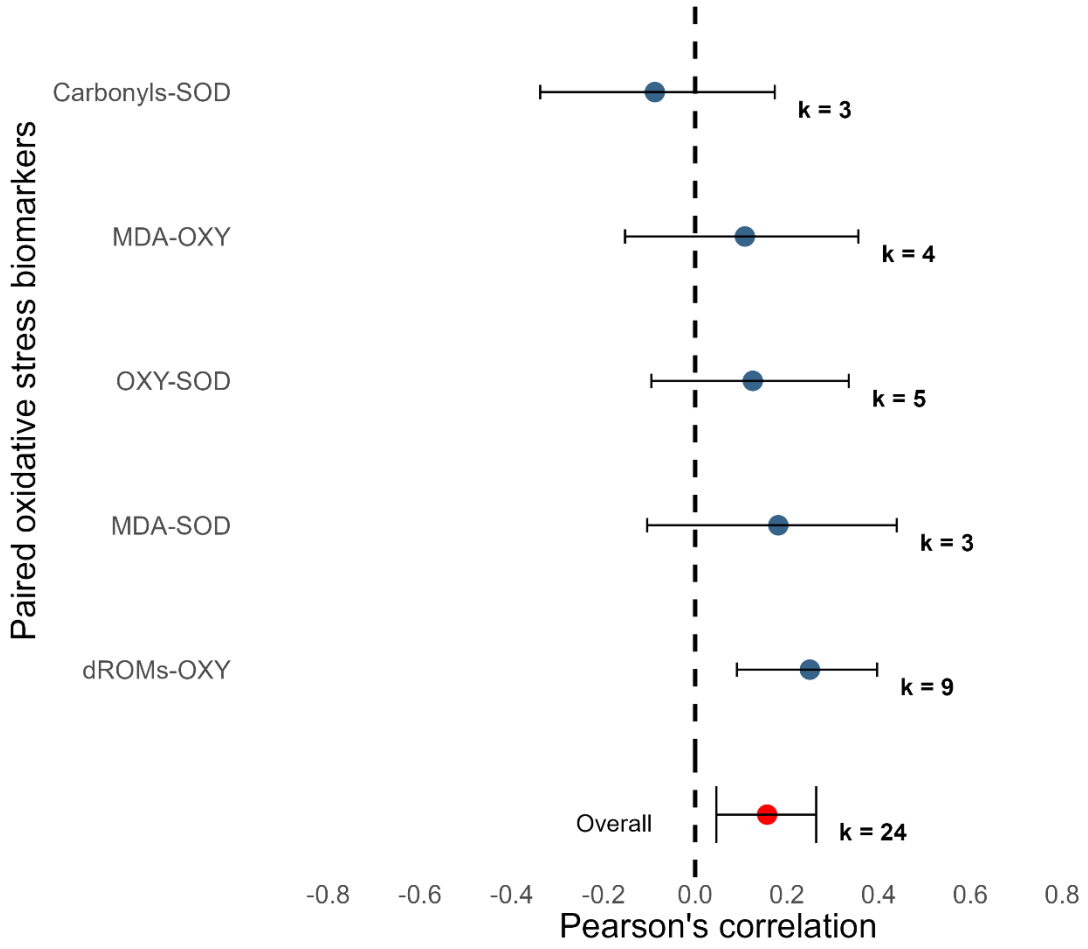

Figure 5: Pearson's correlation for pairs of oxidative stress markers. *The model estimates are shown along with 95% confidence intervals. The dashed line indicates where Pearson's correlation is zero. The paired oxidative stress biomarkers are shown on the y axis and Pearson's correlation (back transformed from Fishers Z) is shown on the x axis. The red point indicates the model estimate for the intercept-only model along with the 95% confidence intervals. "k" represents the number of effect sizes.*

Implications of low repeatability for statistical power: The case of oxidative stress and telomere length

Our meta-analysis revealed that oxidative stress markers exhibit low within-individual repeatability, raising critical concerns about their reliability as biomarkers of physiological condition and fitness. High intra-individual variation reduces the ability to detect biologically meaningful patterns and undermines statistical power. To quantify the implications of this low repeatability, we conducted an individual-based simulation to estimate the sample sizes required to detect a relationship between oxidative stress and telomere length in three biological scenarios: in the first scenario we assume that oxidative stress is fully causal in telomere attrition. The two additional scenarios depict moderate and weak causal effects of oxidative stress on telomere attrition. All simulations were parameterised using biological parameters within the natural range of a hypothetical great tit population (please see supplementary materials).

Our simulations revealed that when oxidative stress repeatability is low (ICC = 0.2), a value consistent with our empirical estimates), sample sizes n > 100 are required to achieve >80% statistical power (Fig. 5A). A previous meta-analysis examining oxidative stress and telomere length correlations (n = 37 studies) found that only 16 studies included more than 100 individuals (13). This power deficit is even more pronounced under scenarios involving moderate or weak true effects of oxidative stress on telomere attrition (Fig. 5A).

Depending on the study system, it may be efficient to increase the statistical power by increasing the number of repeated measurements within individuals as opposed to increasing the sample size (1). Our simulations show that increasing the number of repeated oxidative stress measurements per individual can reduce the total sample size needed to achieve acceptable power (Fig. 5B). Under the scenario in which the true effect of oxidative stress on telomere attrition is moderate, taking three to four repeated measurements per individual reduces the required total sample size by 25-30%, beyond which there are diminishing returns from further increasing the per animal sampling effort. Hence increasing the number of repeated samples can be an effective way of increasing statistical power when the study population is amenable to recaptures. Moreover, such longitudinal studies would provide much needed opportunity for investigating the causes of intraindividual variation. This could ultimately promote developing methods that yield higher conditional repeatability which may be more effective in increasing statistical power.

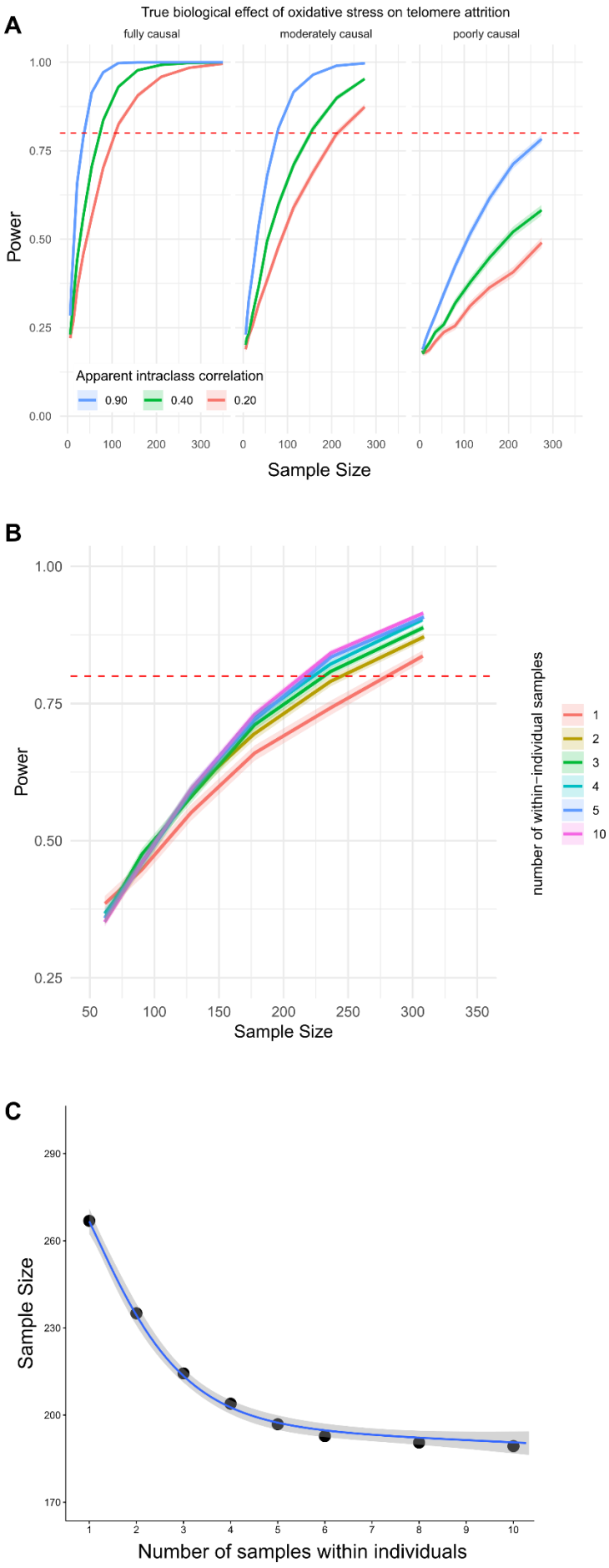

Figure 6: Simulated power of telomere length and oxidative stress correlations against sample size. A) Power is shown for different scenarios of true biological effects of oxidative stress on telomere attrition and for different individual repeatability (intraclass correlations) of oxidative stress. The coloured lines represent the mean value of each repeatability (ICC) group that was simulated. Shaded areas show the 95% confidence intervals. The red dashed line marks the 80% statistical power threshold. Where the coloured line crosses the 80% threshold this would indicate the minimum sample size required to have this level of power to detect the relationship between telomere length and oxidative stress. **B)** Relationship between sample size and repeated measures needed to achieve statistical power above 80% for detecting a correlation between telomere length and oxidative stress, using a scenario where the oxidative stress individual repeatability is low (ICC = 0.2) and causality of the oxidative stress effect on telomere length is moderate. The different colours represent the number of repeated measures per individual. The red dashed line marks the 80% power threshold. Shaded contours reflect the 95% confidence intervals. **C)** The points show the mean sample size required at each number of oxidative stress samples per individual to achieve >80% power to detect the correlation between telomere length and oxidative stress as shown in B. The curve merely illustrates the trend.

General discussion

The heuristic value of a trait as a health biomarker hinges on its ability to capture the links between environmental conditions and fitness. Our meta-analysis reveals that markers of oxidative stress exhibit high intraindividual variation, relative to interindividual variation, implying these markers have low statistical power for capturing individual differences in health and fitness. Furthermore, our moderator analysis indicates that this variability is broadly consistent across different oxidative stress biomarkers, taxa, sexes, and population sources. We have previously argued that strong biomarkers of fitness may be rare, because important fitness traits are likely phenotypically robust and therefore largely invariable with respect to environmental conditions (1). Here we show that the opposite can also be true in the sense that high intraindividual variation induces a 'detection' problem when the source of this variation is unknown, complicating the search for reliable bioindicators. Hence the ideal biomarker appears to comprise of traits that are moderately repeatable, malleable by the environment, and moderately predictive of fitness. Markers of oxidative stress might meet these criteria, but we will need to enhance our understanding of the underlying causes of intraindividual variation, because if we were to unravel these, then this will dramatically enhance the precision with which oxidative stress measurements capture variation in health and fitness among individuals.

Despite the low overall repeatability, several studies included in our meta-analysis reported relatively high repeatability of oxidative stress measurements ($R > 0.6$). What do these studies have in common that may explain their individual repeatability? While it is not possible to make any definite conclusions at this point, we made some tentative observations. First, MDA, particularly when quantified via high-performance liquid chromatography (HPLC), exhibited relatively high repeatability in both wild (2, 3) and controlled settings (4), suggesting its potential as a reliable biomarker of oxidative status. Moderate repeatability was also observed for other biomarkers including vitamin E and SOD under experimental conditions (5, 6), as well as OXY, SOD, and peroxidation index in wild populations (2, 7, 8). In controlled or manipulated conditions, such as dietary interventions or supplementation, high repeatability may reflect reduced environmental noise or greater physiological stability. Conversely, in wild populations, high repeatability tended to occur in individuals experiencing prolonged and predictable physiological demands, such as migration, lactation, and overwinter survival (2, 7, 8). These findings suggest that repeatable oxidative phenotypes may arise when individuals experience sustained metabolic challenges. Notably, high repeatability was detected in both short-term and long-term sampling designs, indicating that repeatability can be high even at extended sampling interval duration. Overall, the lack of consistent repeatability patterns, even within individual studies, highlights the context dependent nature of oxidative stress. This underscores the need for standardised methodologies, and experimental designs that account for ecological and life-history context to better understand individual variation in oxidative stress.

The timescale of within-individual variation in oxidative stress is important. Even if a true biological relationship exists between oxidative stress and fitness, high temporal variability in oxidative stress measurements over short intervals can obscure meaningful links to body condition and fitness. At present, we are not aware of dedicated studies on the fine scaled temporal patterns of within-individual variability of oxidative stress through the life course. However, the stability of oxidative stress has been studied in detail in cell cultures. These in vitro studies show dynamic instability, leading to unpredictable shifts over time, even under standardised conditions (9–12). It remains unclear how well these temporal patterns of oxidative stress in cell cultures that are kept in highly standardized conditions translate to the whole organism. Among the 7 studies (5, 7, 8, 14–17) in our meta-analysis that assessed repeated measurements over short timescales (e.g. days or weeks), most found high variability and low repeatability of oxidative stress markers. Only three of these studies, two of which were under controlled conditions with multiple repeated measurements, reported high repeatability (5, 7, 8). These findings suggest that oxidative stress markers are highly plastic over brief intervals in the majority of cases. To better characterise and understand the causes of this variation, we need longitudinal studies that collect many repeated measurements throughout the life course to obtain fine scale oxidative stress profiles within individuals. This would enable identifying the drivers of intraindividual variation in oxidative stress which would significantly enhance the utility of oxidative stress measurements as a biomarker of body condition and fitness.

Our analysis revealed that correlations among different oxidative stress markers were generally very weak, reinforcing previous findings (38, 53, 54). These poor correlations pose a conceptual challenge because it is generally assumed that different markers of oxidative stress reflect physiological processes that are dynamically linked (18–20). Our findings suggest we need to step away from this

univariate approach as the weak correlations we observed in our meta-analysis indicate that markers of oxidative stress are largely decoupled. The cause of this decoupling could be due to the high intraindividual variability itself: When two traits are poorly repeatable their correlation can be expected to be weak due to the multiplication of poor repeatability scores when correlating the two traits. On the other hand, when two traits dynamically fluctuate in time in a synchronous manner then they could be more strongly correlated than the product of their respective intraclass correlations and we found some evidence for this regarding the correlation between reactive oxygen metabolites and total antioxidant capacity (Fig. 5). These observations call for further research into the question of to what extent different markers of oxidative stress are dynamically linked or whether they may reflect distinct biochemical processes as previously suggested (37, 50, 53, 54). Hence, we need longitudinal multi-marker approaches which remain underutilised in ecological and physiological studies (37, 38).

Finally, the phenomenon of decoupling could be relevant outside the context of oxidative stress. For example, many behavioural traits exhibit high intraindividual variation (20), suggesting that consistent behavioural correlations (i.e., syndromes) may be rare unless behavioural traits are changing across time and contexts synchronously (21). The loss of correlations of suits of traits with age could be indicative of a loss of function or homeostasis, which could be linked to ageing or disease. It will be important to study these multivariate phenotypic changes in relation to fitness to unravel their functional consequences, and oxidative stress may lend itself very well for this purpose.

Materials and Methods

Literature search

We developed a search string using keywords derived from a review of the oxidative stress literature as follows:

(“Oxidative stress” OR "Antioxidant capacity" OR “Antioxidant*” OR "Oxidative damage") AND (“Repeated*” OR “Repeated measure” OR “Repeatability*” OR “Longitudinal*” OR “Longitudinal study” OR “Within Individual” OR “Within Individual study”) AND (“Wildlife*” OR “Wild animal” OR “Wild population” OR “Wild*” OR “Bird*” OR “Avian*” OR “Invertebrate*” OR “Fish*” OR “Herpetofauna*” OR “Vertebrate*” OR “Mammal*”) NOT (“Human*” OR “Woman*” OR “Women*” OR “Man*” OR “Children*” OR “Patients*” OR “People*” OR “Men*” OR “Child*”)

This search string was used to collect literature from both the Web of Science core collection database (last search 7th of December 2023) and Scopus database (last search 4th of June 2024). After removing duplicates, we recorded the number of results for each iteration following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) framework. The PRISMA diagram summarising the selection process is shown in Figure S2. We were able to include effect sizes from 22 studies (2, 4–8, 13–16, 22–33).

Inclusion criteria

To be included in the meta-analysis, studies had to meet the following criteria:

1) Measured at least one biomarker of oxidative stress (oxidative damage or antioxidant defence).
2) Took at least two repeated oxidative stress measurements from the same individual, with the time interval between measurements reported.
3) Focused on non-human organisms.
4) Provided access to raw individual-level data, either publicly available or obtained via author correspondence.

We included both experimental and observational studies across various environments to ensure broad ecological and taxonomic representation.

Screening process

We used a two-step screening process to identify eligible studies. First, titles and abstracts were screened using the “MetaGear” (34) package in R (35) to identify studies involving oxidative stress in non-human organisms. Because repeat sampling was not always explicitly stated in abstracts, we next screened the full texts of selected studies, verifying in the methods sections that repeated measurements on the same individuals were conducted.

Data extraction

From each eligible study, we extracted biological and methodological information, including species, taxonomic class, sample size, number of samples per individual (mean), time between repeated samples (mean), oxidative stress biomarkers measured, environment, and study type (manipulated vs. non-manipulated). For experimental studies, we separated treatment and control groups, and calculated repeatability estimates independently for each group. To ensure comparability across study types, control groups from experimental studies were pooled with non-experimental studies and were referred to as non-manipulated, whereas treatment groups within experimental studies were referred to as manipulated.

Environment was categorised as wild, captive, wild captive (wild individuals brought into captivity for the purposes of the study), or livestock (Animals used for agricultural purposes, including horses, cattle and sheep). Biomarkers were classified based on whether they measured oxidative damage (lipid peroxidation, protein damage, DNA damage, or nonspecific damage) or antioxidant defences (enzymatic or non-enzymatic) (see Table S5).

Repeatability estimation

To calculate within-individual repeatability of oxidative stress measurements, we obtained the raw data from each study. For studies where raw data was not directly available, we contacted the authors (Table S4). We estimated repeatability by determining the intraclass correlation coefficient (ICC) using ANOVAs based on equations 4 and 5 from Nakagawa & Schielzeth (2010) (36), along with the standard error and 95% confidence intervals. This ANOVA-based method is recommended when repeatability is very low (<0.05) which was often the case in our dataset, as the linear mixed model (LMM) approach tends to overestimate low repeatability values (36, 37). While we used the ANOVA-based approach for our main analyses, we confirmed the consistency of the results with LMM-derived estimates (Fig. S3). Importantly, rather than calculating conditional repeatability by including covariates within the repeatability model itself, we stratified the data by relevant variables, specifically by sex, oxidative stress biomarker, and study type (manipulation vs. non-manipulation) to calculate separate ICC values for each subset. This approach allowed us to subsequently use these variables as moderators in the meta-analysis to explore potential sources of variation in repeatability. This yielded 123 effect sizes from 22 species across two classes from 22 studies that were published between 2013 and 2023 (species listed in Table S6).

Oxidative biomarker correlations.

For studies measuring multiple oxidative stress biomarkers within the same individuals, we calculated Pearson’s correlation coefficients between biomarker pairs using the cor function in R. These values were included in an additional meta-analysis to assess the consistency of biomarker co-variation across studies.

Meta-analysis approach.

All analysis were formed in R version 4.3.3 (76).

All meta-analyses were conducted using a multi-level meta-analysis structure with the metafor package (38). ICC and Pearson’s r values were standardised using Fishers Z transformation (37) and then were back transformed for interpretation in figures and tables. Corresponding sampling variances were calculated using the escalc function.

We first estimated the overall ICC using an intercept-only random effects model. We then tested whether variation in repeatability could be explained by moderators including study type, environment, taxonomic class, oxidative stress biomarker, mean time between measurements (log-transformed),

and the average number of samples per individual (log-transformed). Subgroup analysis further compared between oxidative damage and antioxidant defence categories using the respective biomarker type as a moderator in each analysis (Table S4).

The meta-analysis model exploring biomarker correlations used the same model structure, with biomarker pair identity as a categorical moderator.

All models included study ID and observation ID as random effects to account for non-independence and hierarchical structure (39, 40). Total relative heterogeneity and the contribution of each random effect was quantified using the i2_ml function from the OrchaRd package (41).

Lastly, we evaluated potential publication bias by fitting two multilevel meta-regression models, with the square root of inverse sampling variance and mean-centred year of publication as moderators to test for small study effects and time lag bias respectively (Fig S4) (42, 43).

The methods for the simulation can be found in the supplementary materials, Appendix two.

Conditional repeatability comparison

To evaluate whether controlling for biological covariates alters repeatability estimates of oxidative stress biomarkers, we performed an additional meta-analysis using a subset of studies (n = 8) for which the raw data included information about age, and in some cases, sex and treatment group. For each oxidative stress biomarker in these studies, we estimated repeatability under four model conditions: (1) unadjusted (raw, no covariates), (2) age adjusted, (3) age and sex adjusted, and (4) age and treatment group adjusted.

Repeatability was calculated using the rptR (36, 37) package in R, employing an LMM-based estimation approach. This method allows for covariates to be included as fixed effects within the model, providing estimates of conditional repeatability. We then performed a meta-analysis using repeatability model type as a moderator to assess whether controlling for these covariates influenced repeatability estimates.

Availability of data and materials

The code and data used in this study will be made available online following peer review at the time when the manuscript is accepted for publication

Funding

This project was funded through a grant from Wild Animal Initiative to Davide M Dominoni. Jelle Boonekamp was supported through a grant from the Leverhulme Trust (RPG-2024-207).

Acknowledgments

Thanks to all of the authors who responded to our data enquiries, including Quentin Schull, Sarah Guindre-Parker, Helga Sauerwein, Melissah Rowe, Brenna Gormally, Aurora Ramierez-Perez, Mireia Blanco, Shelia Stivanin, Valeria Marasco, and Martin Tesicky.

Author contributions

Rachel Reid, Davide M Dominoni and Jelle Boonekamp conceived of and designed the study. Rachel Reid and Jelle Boonekamp developed the simulation. Rachel Reid ran the meta-analyses and drafted the manuscript with help from all co-authors. All authors read and approved the final version.

References

1. M. A. Goulet, D. Cousineau, The Power of Replicated Measures to Increase Statistical Power. *Adv Methods Pract Psychol Sci* **2**, 199–213 (2019).

2. J. J. Boonekamp, R. Dijkstra, C. Dijkstra, S. Verhulst, Canalization of development reduces the utility of traits as fitness biomarkers: feather fault bars in nestling birds. *Funct Ecol* **31**, 719–727 (2017).

3. T. W. Bodey, *et al.*, Consistent measures of oxidative balance predict survival but not reproduction in a long-distance migrant. *Journal of Animal Ecology* **89**, 1872–1882 (2020).

4. L. L. Christensen, *et al.*, Marker-dependent associations among oxidative stress, growth and survival during early life in a wild mammal. *Proceedings of the Royal Society B: Biological Sciences* **283** (2016).

5. K. G. Orquera-Arguero, I. Casasús, J. Ferrer, M. Blanco, Beef cows' performance and metabolic response to short nutritional challenges in different months of lactation. *Res Vet Sci* **159**, 26–34 (2023).

6. V. Marasco, *et al.*, Environmental conditions can modulate the links among oxidative stress, age, and longevity. *Mech Ageing Dev* **164**, 100–107 (2017).

7. E. Velázquez-Cantón, *et al.*, Effect of Selenium and Vitamin E Supplementation on Lactate, Cortisol, and Malondialdehyde in Horses Undergoing Moderate Exercise in a Polluted Environment. *J Equine Vet Sci* **69**, 136–144 (2018).

8. C. Eikenaar, E. Winslott, S. Hessler, C. Isaksson, Oxidative damage to lipids is rapidly reduced during migratory stopovers. *Funct Ecol* **34**, 1215–1222 (2020).

9. M. Těšický, *et al.*, Longitudinal evidence for immunosenescence and inflammaging in free-living great tits. *Exp Gerontol* **154** (2021).

10. J. Musakhanian, J. D. Rodier, M. Dave, Oxidative Stability in Lipid Formulations: a Review of the Mechanisms, Drivers, and Inhibitors of Oxidation. *AAPS PharmSciTech* [Preprint] (2022).

11. M. Ayuso, S. Van Cruchten, C. Van Ginneken, A medium-throughput system for in vitro oxidative stress assessment in IPEC-J2 cells. *Int J Mol Sci* **21**, 1–18 (2020).

12. C. S. Shiva Shankar Reddy, M. V. V. Subramanyam, R. Vani, S. Asha Devi, In vitro models of oxidative stress in rat erythrocytes: Effect of antioxidant supplements. *Toxicology in Vitro* [Preprint] (2007).

13. B. Halliwell, Oxidative stress in cell culture: An under-appreciated problem? *FEBS Lett* **540**, 3–6 (2003).

14. B. M. G. Gormally, R. Estrada, H. Yin, L. M. Romero, Recovery from repeated stressors: Physiology and behavior are affected on different timescales in house sparrows. *Gen Comp Endocrinol* **282** (2019).

15. G. J. Kolling, *et al.*, Milk production and hematological and antioxidant profiles of dairy cows supplemented with oregano and green tea extracts as feed additives. *Revista Brasileira de Zootecnia* **51** (2022).

16. J. C. Noguera, S.-Y. Kim, A. Velando, Family-transmitted stress in a wild bird. *PNAS* **114**, 1–6 (2017).

17. H. Sauerwein, *et al.*, Acute phase proteins and markers of oxidative status in water buffalos during the transition from late pregnancy to early lactation. *Vet Immunol Immunopathol* **228** (2020).

18. M. M. Cortese-Krott, *et al.*, The Reactive Species Interactome: Evolutionary Emergence, Biological Significance, and Opportunities for Redox Metabolomics and Personalized Medicine. *Antioxid Redox Signal* [Preprint] (2017).

19. L. L. Ji, D. Yeo, Oxidative stress: an evolving definition. *Fac Rev* **10** (2021).

20. V. I. Lushchak, Free radicals, reactive oxygen species, oxidative stress and its classification. *Chem Biol Interact* [Preprint] (2014).

21. A. M. Bell, S. J. Hankison, K. L. Laskowski, The repeatability of behaviour: a meta-analysis. *Anim Behav* [Preprint] (2009).

22. G. Polverino, K. M. Buchholz, C. T. Goulet, M. Michelangeli, D. G. Chapple, Temporal repeatability of behaviour in a lizard: implications for behavioural syndrome studies. *Evol Ecol* **37**, 401–418 (2023).

23. C. Récapet, M. Arrivé, B. Doligez, P. Bize, Antioxidant capacity is repeatable across years but does not consistently correlate with a marker of peroxidation in a free-living passerine bird. *J Comp Physiol B* **189**, 283–298 (2019).

24. L. L. Christensen, *et al.*, Plasma markers of oxidative stress are uncorrelated in a wild mammal. *Ecol Evol* **5**, 5096–5108 (2015).

25. M. M. Skrip, S. R. Mcwilliams, Oxidative balance in birds: An atoms-to-organisms-to-ecology primer for ornithologists. *J Field Ornithol* **87**, 1–20 (2016).

26. K. A. Herborn, *et al.*, Age, oxidative stress exposure and fitness in a long-lived seabird. *Funct Ecol* **30**, 913–921 (2016).

27. M. Rowe, K. L. Pierson, K. J. McGraw, Exploratory behavior is associated with plasma carotenoid accumulation in two congeneric species of waterfowl. *Behavioural Processes* **115**, 181–190 (2015).

28. D. L. Cram, J. D. Blount, A. J. Young, The oxidative costs of reproduction are group-size dependent in a wild cooperative breeder. *Proceedings of the Royal Society B: Biological Sciences* **282** (2015).

29. E. I. K. Vitikainen, *et al.*, Evidence of oxidative shielding of offspring in a wild mammal. *Front Ecol Evol* **4** (2016).

30. S. Guindre-Parker, D. R. Rubenstein, No short-term physiological costs of offspring care in a cooperatively breeding bird. *Journal of Experimental Biology* **221** (2018).

31. S. Guindre-Parker, *et al.*, The oxidative costs of territory quality and offspring provisioning. *J Evol Biol* **26**, 2558–2565 (2013).

32. Q. Schull, *et al.*, The oxidative debt of fasting: Evidence for short- to medium-term costs of advanced fasting in adult king penguins. *Journal of Experimental Biology* **219**, 3284–3293 (2016).

33. V. A. Viblanc, *et al.*, Maternal oxidative stress and reproduction: Testing the constraint, cost and shielding hypotheses in a wild mammal. *Funct Ecol* **32**, 722–735 (2018).

34. M. Hau, *et al.*, Repeated stressors in adulthood increase the rate of biological ageing. *Front Zool* **12** (2015).

35. M. J. Lajeunesse, Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods Ecol Evol* **7**, 323–330 (2016).

36. R Core Team, R: A language and environment for statistical computing . [Preprint] (2024).

37. S. Nakagawa, H. Schielzeth, Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews* [Preprint] (2010).

38. B. Holtmann, M. Lagisz, S. Nakagawa, Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: a meta-analysis. *Funct Ecol* **31**, 685–696 (2017).

39. W. Viechtbauer, Conducting meta-analyses in R with the metafor. *J Stat Softw* **36**, 1–48 (2010).

40. B. Fernández-Castilla, *et al.*, The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behav Res Methods* **52**, 2031–2052 (2020).

41. J. R. Dettori, D. C. Norvell, J. R. Chapman, Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider. *Global Spine J* **12**, 1624–1626 (2022).

42. S. Nakagawa, *et al.*, orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. *Methods Ecol Evol* **2023**, 2003–2010 (2023).

43. J. Koricheva, E. Kulinskaya, Temporal Instability of Evidence Base: A Threat to Policy Making? *Trends Ecol Evol* **34**, 895–902 (2019).

44. S. Nakagawa, *et al.*, Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods Ecol Evol* **13**, 4–21 (2022).