EDA: Oxidative Stress, Inflammation, and Glucocorticoid Stress

Daniel J. Naumenko

2022-09-14

Project Description

This report is part of a straightforward project intended to determine whether immune and stress responses cause/result in/associate with oxidative stress in a wild orangutan population. Some or all of these findings may be merged with the tuanan_dietary_restriction_oxidative_stress_inflammation_cortisol project focusing on DR. However, I want to work out this biomarker-only part first, and then relate them to diet, and see if the two stories should be merged or if the amount of information is too substantial to put them together.

The question is meta-analyzed in birds in Costantini and Moller. I will be focusing on oxidative stress (acronym: OS) as a response, with immune responses (WBC, neopterin, cytokines) and stress responses (cortisol) as potential drivers of increased oxidative stress. As part of this, I would want to test whether there are age-sex differences in oxidative stress, and if I can, try to detect seasonal trends perhaps using the GAMM approach of polansky and robbins 2013 or by one of Gavin Simpson's blogposts.

I may want to include the creatinine-residual data from caitlin, as there are myogenic and myoanabolic links between oxidative stress and lean muscle mass. ROS can stimulate expression and activity of skeletal muscle protein degradation pathways (muscle wasting).

Co-Authors: Erin R. Vogel, Rebecca S.A. Brittain, Tim Bransford (? cortisol), Alysse Moldawer (? cytokines), Lync Moldawer (? cytokines), Sri Suci Utami Atmoko, Caitlin S.A. O'Connell (ran a lot of neopterin).

How do orang neopterin levels compare to neopterin in other NHPs? - Probably answered in Liz's dissertation.

Data Details

- dimensions
- for all urine samples
  - Number of samples per orangutan
  - number of samples per field worker
  - number of samples per fai tercile
  - number of samples per year
- biomarker data inclusion criteria
  - start with total number of samples with each measurement
  - remove poor CVs
  - remove low specific gravity
  - maybe remove 1.5 times IQR outliers, but definitely box plot before this step
  - non-first morning voids

Table 1: Table 1a: All assayed samples without any removed and no specific gravity correction applied

biomarker	non_na	minimum	maximum
8-OHdG	742	0.102	255.544
Cortisol	995	152.948	4405531.000
G-CSF	460	0.000	12864.000
IL-10	457	0.000	5383.000
IL-1Ra	452	0.000	507.750
IL-8/CXCL-8	459	0.000	2751.000
MCP-1	461	0.000	919.540
Neopterin	1165	1.560	14243.201
TNF_a	461	0.000	919.540
Total Antioxidant Capacity	554	0.000	36.538

  - contaminated with fecal matter
  - status not "ok"

- - Maria file from caitlin paper with individual IDs

To Check

- orangutan names, make sure they're correct
- double check age-classes
- if a cytokine is marked as zero, it was below the limit of detection
  - i say give it a random non-zero number that is within the LOD range

I still need to do something with the zero values for the cytokines, as they're not really zero.

How many biomarker values does each individual orangutan have? I need to double check that the values in the database under Orangutan_ID are indeed correct. Maria sent an excel file with a list of follow and sample numbers and the correct names and age_sex classes. I should merge this in and use those instead of any values in the urine database because I am not sure they've been QC'd.

This table has been arranged by number of 8-OHdG samples in descending order. Total biomarker values indicates the sum of all assayed samples. I also only used samples that satisfied primary inclusion criteria: kept in thermos in field, SG not dilute, a first morning void, and the CV for the given assay sample was below 15%.

How many assayed samples of each biomarker meet each inclusion criterion? I want to look at each of my inclusion criterion to determine whether any lead to significant loss of sample numbers or change balance fo the data, ultimately reducing power. When I know that a given exclusion criterion reduces power, I can make a judgement call as to what to do.

The inclusion criteria are as follows:

1. Sample is not dilute (specific gravity > 1.003)
2. Sample assay was precise (CV is equal to or below 15%)
3. Sample is from a first morning void

Table 2: Table 1c: All assayed samples by individual orangutan

Orangutan_ID	8-OHdG	Cortisol	G-CSF	IL-10	IL-1Ra	IL-8/CXCL-8	MCP-1	Neopterin	TN
Mindy	64	56	15	14	15	17	16	62	
Milo	49	37	11	10	9	10	8	60	
Juni	44	48	18	16	14	18	15	55	
Dayak	35	25	5	4	4	6	4	32	
Kerry	35	23	8	9	9	10	8	42	
Kondor	25	26	2	3	2	2	2	30	
Niko	25	29	6	6	5	4	4	31	
Tomi	25	25	5	5	5	5	3	28	
Jinak	24	31	12	11	10	11	11	23	
Inul	22	20	6	5	5	6	6	21	
Wodan	21	25	8	7	7	10	8	29	
Mawas	19	12	0	1	2	2	2	29	
Teju	13	9	0	0	1	0	1	11	
Dado	12	0	0	0	0	0	0	5	
Desy	12	5	4	4	4	4	4	16	
Henk	12	15	8	8	10	10	10	16	
Otto	12	15	4	4	5	5	5	14	
Sidony	12	10	13	13	13	13	13	15	
Ekko	11	6	1	1	1	1	1	17	
Gismo	11	5	1	1	1	1	1	17	
Helium	10	10	5	5	5	5	5	13	
Sugus	9	8	0	1	1	1	1	13	
Ted	9	7	0	0	0	0	0	10	
Jip	8	7	3	3	2	3	3	8	
Cinta	7	7	0	0	0	0	0	7	
Ipsy	6	6	1	1	1	1	1	9	
Pinky	6	12	0	0	0	0	0	16	
Unknown unflanged male	6	3	1	1	1	1	1	5	
Bobo	5	6	0	1	1	1	1	4	
Tina	5	7	0	0	0	0	0	10	
Vini	5	0	0	0	0	0	0	4	
Katmandun	4	4	3	3	4	3	3	2	
Sony	4	0	1	2	2	3	2	3	
Streisel	4	4	1	0	1	1	1	4	
Tarzan	4	5	3	3	3	3	2	5	
Charlie	3	2	0	0	0	0	0	3	
Kino	3	3	1	1	1	1	1	3	
Chaz	2	3	4	4	4	4	4	2	
Kiki	2	2	0	0	0	0	0	3	
Max	2	0	0	0	0	0	0	4	
Unknown flanged male	2	0	0	0	0	0	0	6	
Whisky	2	0	0	0	0	0	0	2	
Arthur	1	1	0	0	0	0	0	4	
Dolay	1	1	1	1	1	1	1	1	
Fugit	1	4	0	0	0	0	0	5	
Guapo	1	1	0	0	0	0	0	1	
Merkur	1	0	0	0	0	0	0	1	
MindyMawas	1	1	1	1	1	1	1	0	
Momo	1	2	0	0	0	0	0	1	
Preman flanged	1	0	0	0	0	0	0	0	
Tom	1	1	0	0	0	0	0	1	
Tuk-Tuk	1	1	3 0	0	0	0	0	1	
Unknown adolescent female	1	0	0	0	0	0	0	1	
Wilma	1	1	0	0	0	0	0	1	
Zola	1	1	0	0	0	0	0	1	

Table 3: Table 1d: The number of assayed samples by biomarker that meet each inclusion criterion

Biomarker	# of Total Assayed Samples	SG > 1.003	Stored in Thermos	First Morning Voids
8-OHdG	742	740 (99.73%)	715 (96.36%)	649 (87.47%)
Cortisol	995	932 (93.67%)	806 (81.01%)	685 (68.84%)
G-CSF	460	384 (83.48%)	291 (63.26%)	276 (60%)
IL-10	457	384 (84.03%)	287 (62.8%)	273 (59.74%)
IL-1Ra	452	376 (83.19%)	281 (62.17%)	266 (58.85%)
IL-8/CXCL-8	459	383 (83.44%)	289 (62.96%)	273 (59.48%)
MCP-1	461	385 (83.51%)	290 (62.91%)	275 (59.65%)
Neopterin	1165	1153 (98.97%)	1043 (89.53%)	846 (72.62%)
TNF_a	461	386 (83.73%)	291 (63.12%)	275 (59.65%)
Total Antioxidant Capacity	554	552 (99.64%)	537 (96.93%)	460 (83.03%)

4. Assay was determined to have run appropriately (Sample assay status is OK, not set to 'rerun')

Ok, I will calculate percentages for each of these variables, surround the percentages in parentheses and add a percent sign, then merge the pairs of columns together in the final product. Should be able to use column_spec to add the parentheses and percentages.

After filtering out samples that were not stored in a cold thermos in the field, were not first morning voids, were too dilute, and had poor CVs, the following number of samples remained for each biomarker.

604, 446, 750, 543, 152, 149, 150, 164, 149, 174

I could produce boxplots of each of these inclusion categories to see how their exclusion changes the spread and central tendency of each biomarker.

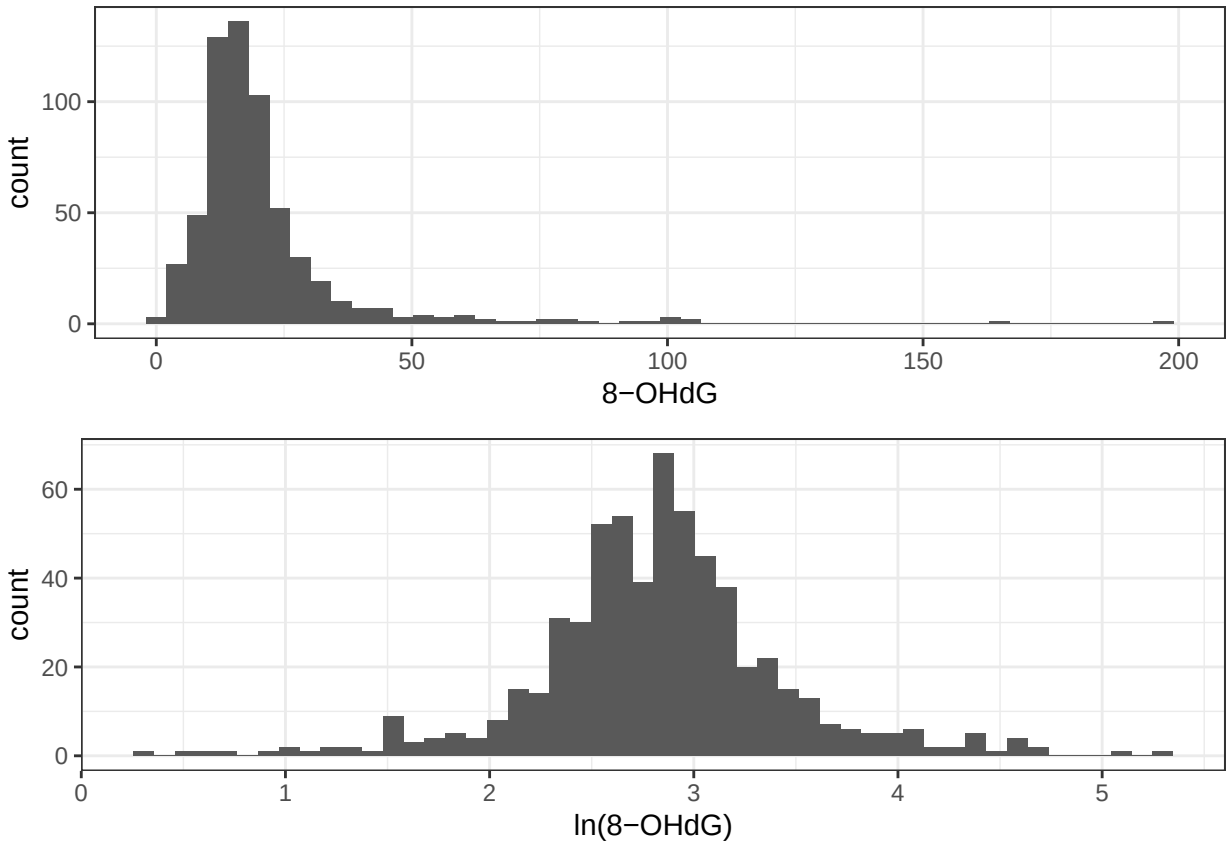
How many samples have a status of "ok"? For this one I need to make sure all records have a value here. Otherwise some NAs may get removed. I have not yet created this table.

Variation

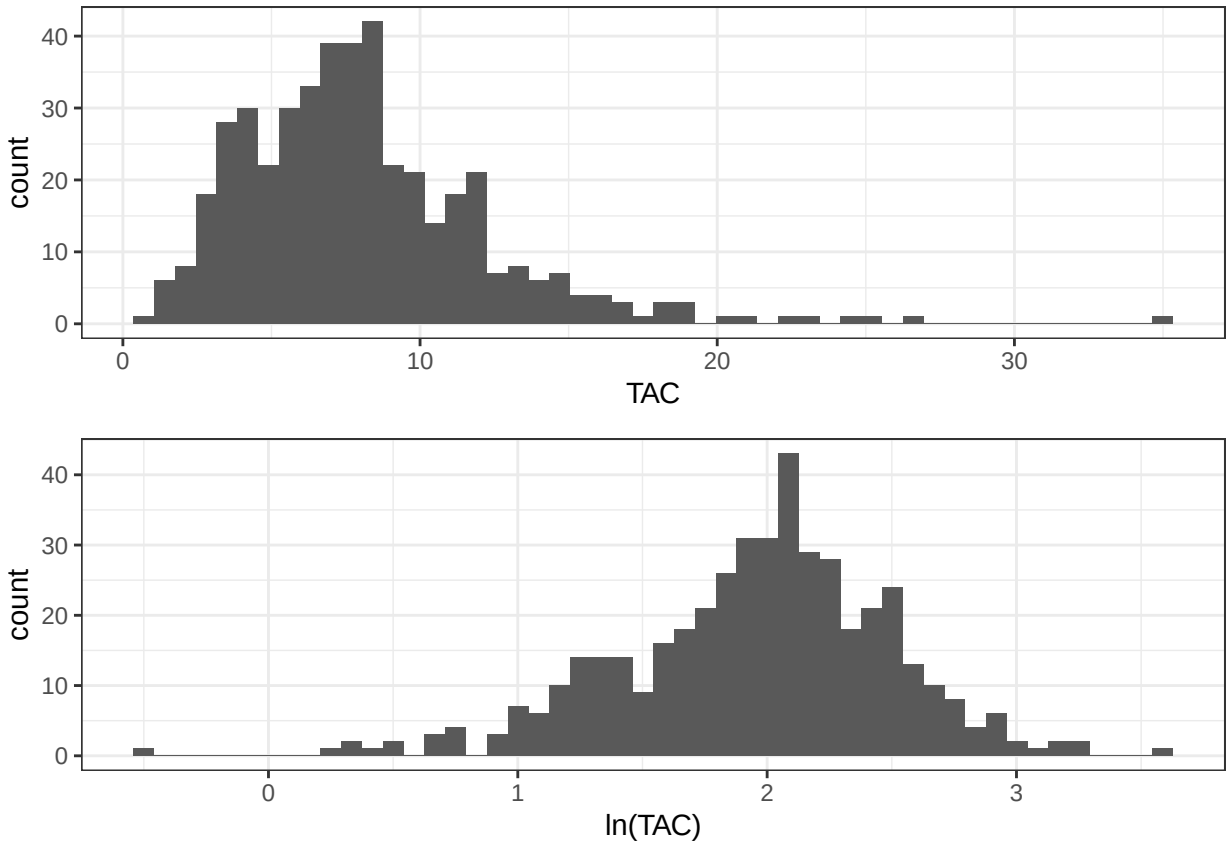
General Notes: I am not yet sure how I want to deal with outliers. Many papers use the 1.5 * IQR method; this is reasonable. The outliers themselves should be investigated to see if I can determine why they are outliers. If there is an error, I can remove them. If they are part of natural variation, I may keep or remove them. If they are removed but seem to be accurate values, I may want to provide supplementary information on the samples in a table providing any relevant information on them.

8-OHdG Variation There appears to be a few pretty large outliers in the >90 range. A natural log transformation creates a vaguely normal distribution with some pretty long tails on either end and the potential for a little bimodality around ln2.7. The skewness of the SG-corrected 8-OHdG is 4.4674973 and kurtosis is 33.2902414

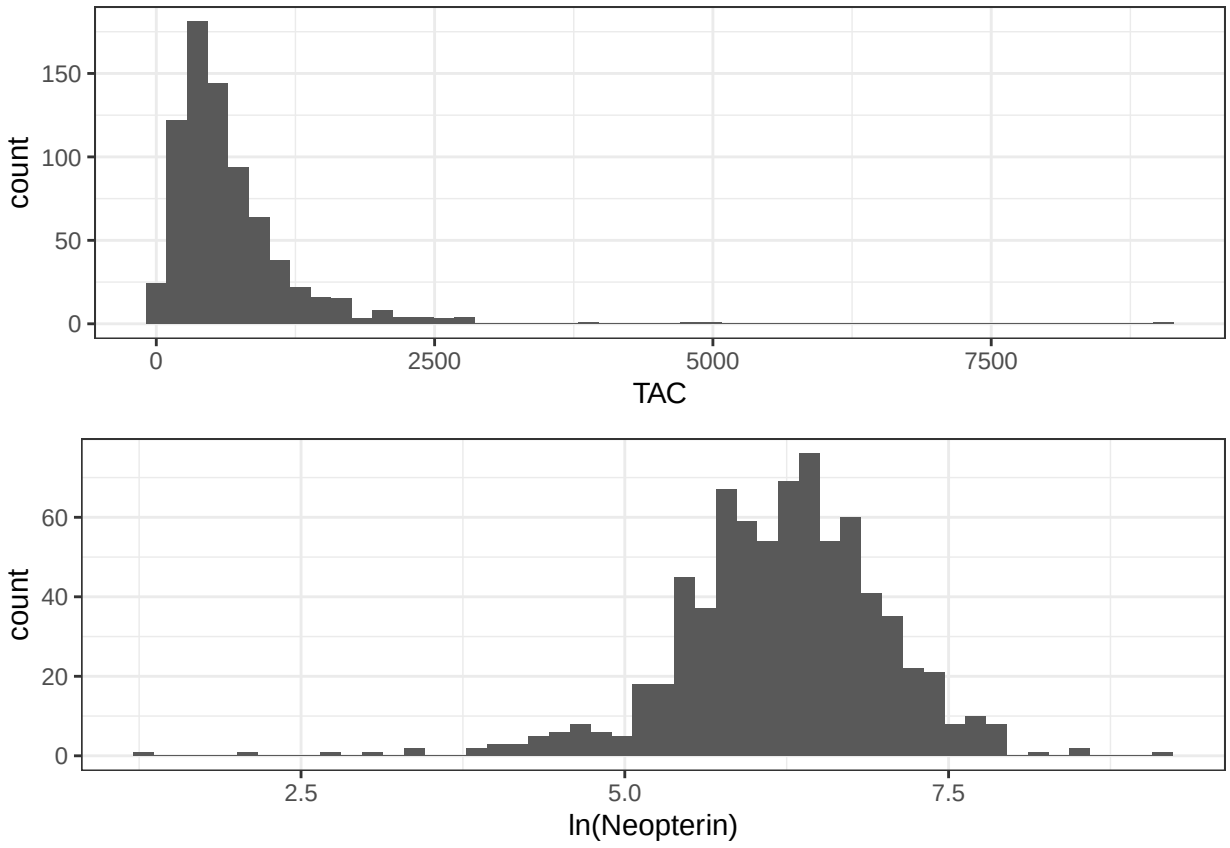
The skewness of the natural logged 8-OHdG is -0.0017255 and the kurtosis is 5.4216779. So there is a considerable reduction in both skew and kurtosis after a natural log transformation, but the distribution of these data should be explored further. Is there a reason for that near-bimodality? Age-sex class?



TAC Variation Variation in TAC seems to be the most interesting at first glance, as it is the most unique distribution. The units are in uM, not umol/L. But the transformation is a linear scale so the shape of the distribution will not change. It is possible that removing TAC values < 20uM might help, but the right tail is not too extreme. I am also reluctant to remove TAC values because they are so essential to interpreting oxidative stress. These high TAC values may be staving off (or failing to stave off) oxidative stress. If 8-OHdG is low during these high periods, it suggests the individual is dealing with something that is causing a lot of oxidative stress, but the body is otherwise successfully dealing with the problem as DNA damage does not increase. On the otherhand, if 8-OHdG is high in this upper range of TAC, it suggests the body is mounting a strong defense but it is not enough. I would want to check this. In addition, I want to see which individuals these high values are from, and if they have lower values as well or if some individuals just tend to produce higher values. Plot TAC over time by individual and see if any are typically high at all times. If such an individual exists, it is probably part of their own natural biology, and not indicative of chronic free radical generation.



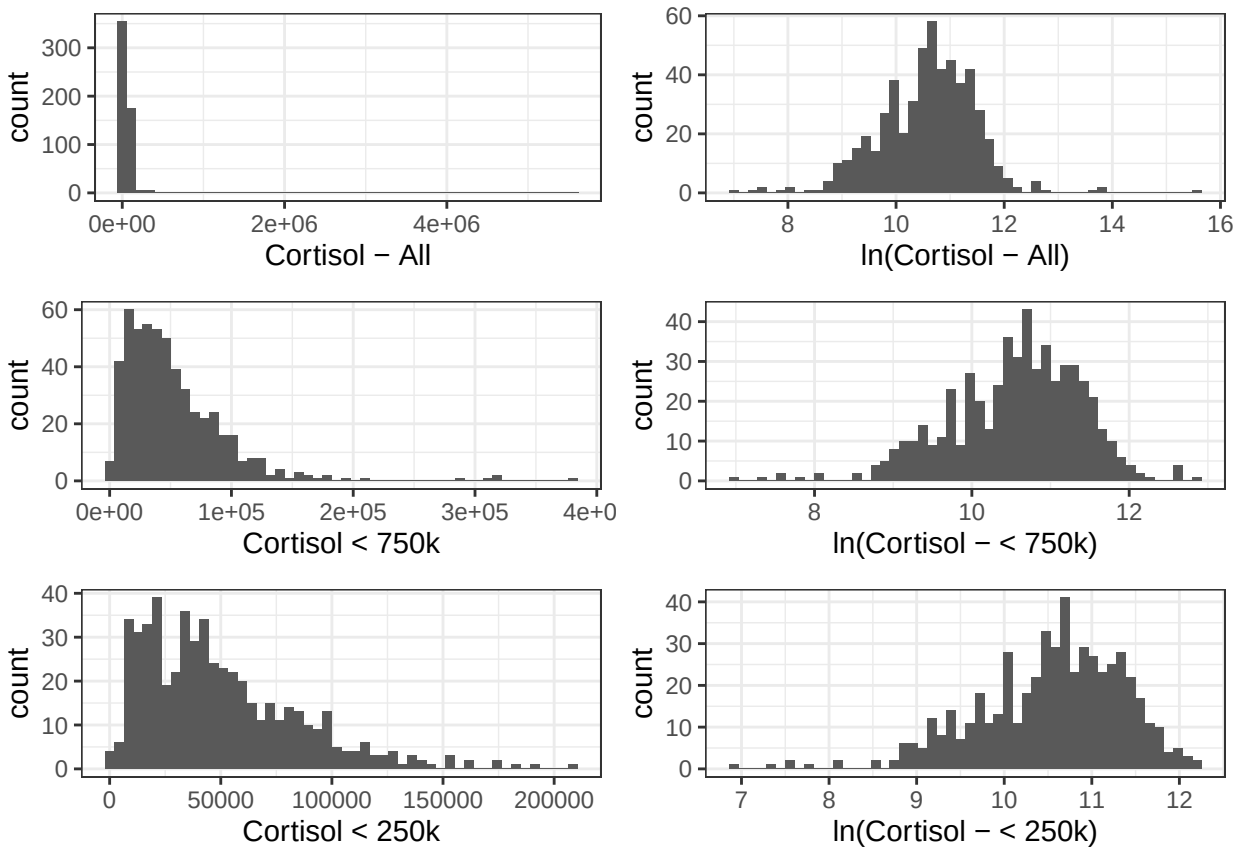
Neopterin Variation Neopterin, like cortisol below, has some pretty extreme outliers, though not to the same extent as cortisol. A natural log transformation does seem to bring these values together. However, I think I'd rather model the actual data using GAMMs as those values are much more interpretable than trying to understand the logged effects of biomarkers on each other.



Cortisol Variation You can see there are a small handful of extraordinarily extreme values of cortisol. They definitely need to be checked to determine if they are indeed correct. To temporarily view the distribution without these extreme values, I replicated the plots with X removed.

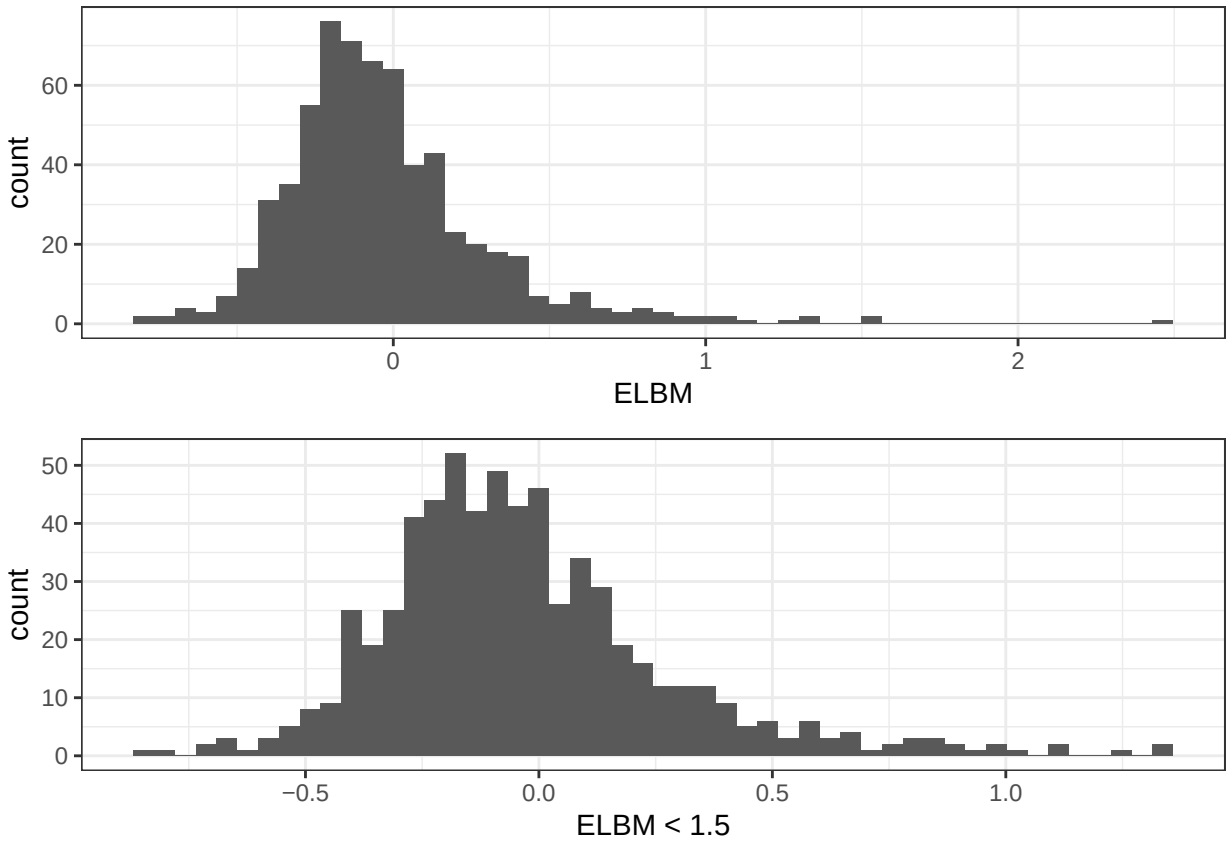
Table 4: Table: Skewness and Kurtosis of cortisol at different levels of outlier removal; only samples that meet inclusion criteria used

	Skewness	Kurtosis
All cortisol values	19.902720	432.612382
Cortisol $<$ 750k	2.758784	16.210378
Cortisol $<$ 250k	1.205298	4.656539



There is a pretty dramatic reduction in kurtosis when the most extreme outliers are removed. The reasoning for these outliers, be they error or some latent features, should be explored. In-depth exploration is probably beyond the scope of this paper; some decision must be made for which cortisol values to include and this decision must be justifiable.

ELBM Distribution I think I will keep in the outliers for now, and then see what to do about them later. Later note: as I was plotting covariation below, I routinely removed the ELBM outlier above 2.



General Notes on Biomarker Distribution There is some pretty heavy right tails in these data, regardless of how many outliers are removed. Special attention will need to be paid to these outliers given their sheer scale.

There is an outlier-robust GAM fitting package (rgam) that may be of use. It uses a backfitting algorithm with weights derived from robust quasi-likelihood equations. Original publication: Azadeh, A. and Salibian-Barrera, M. (2011). *An outlier-robust fit for Generalized Additive Models with applications to disease outbreak detection.*

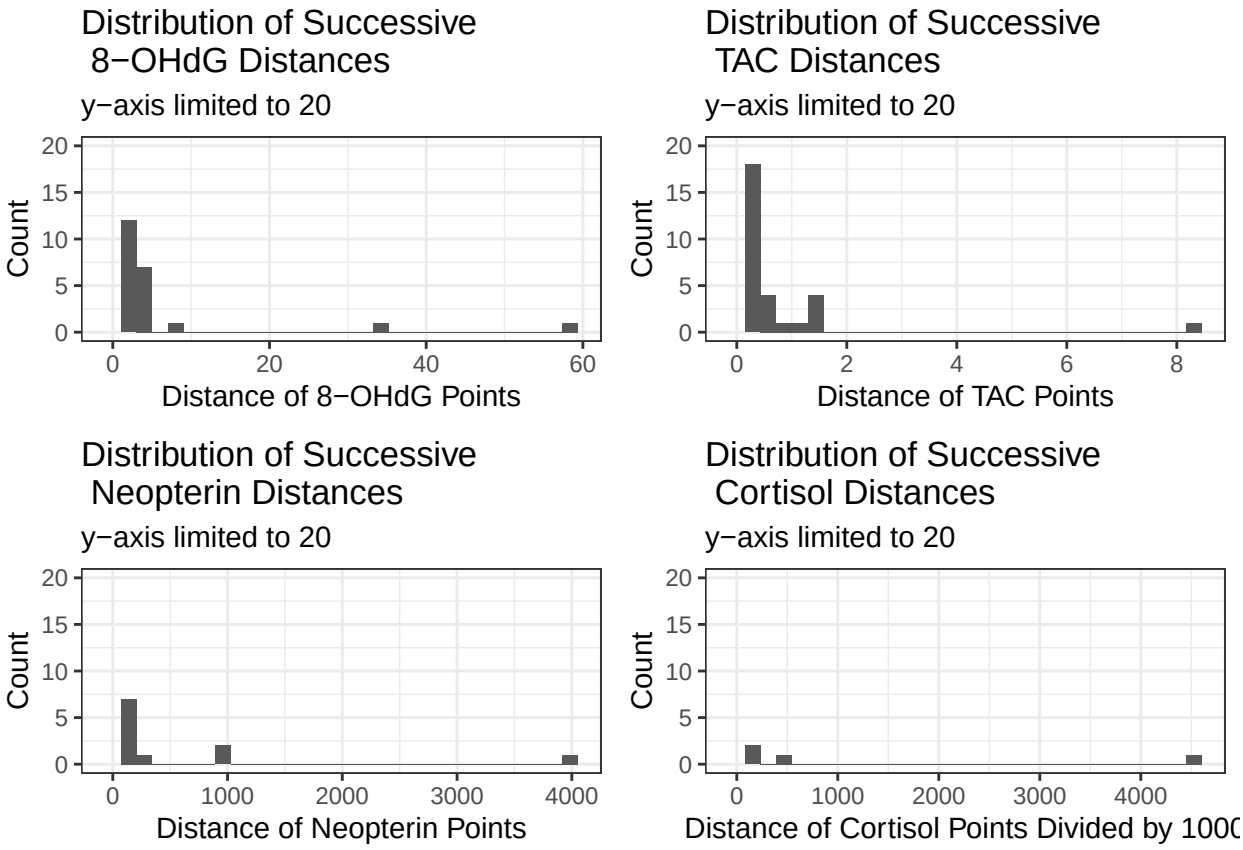
Outlier Analysis

Before proceeding beyond my analysis of variation of each of the biomarkers, I need to better understand what data points are outliers, why they are so different from the rest of the data, and then decide how to handle them. There are a range of techniques I could use to detect filters: z-score and flag points beyond the -3:3 z-score range, flag points outside the 1.5 * IQR range, Expectation Maximization, Mahalanobis Distance, and DBSCAN. I will focus on the last 3 first as ML algorithms and compare their findings. I will likely stick with DBSCAN as it requires fewer assumptions, but I want to check first.

I want to flag each biomarker as an outlier under each method so that I can then determine if there is anything going on about them.

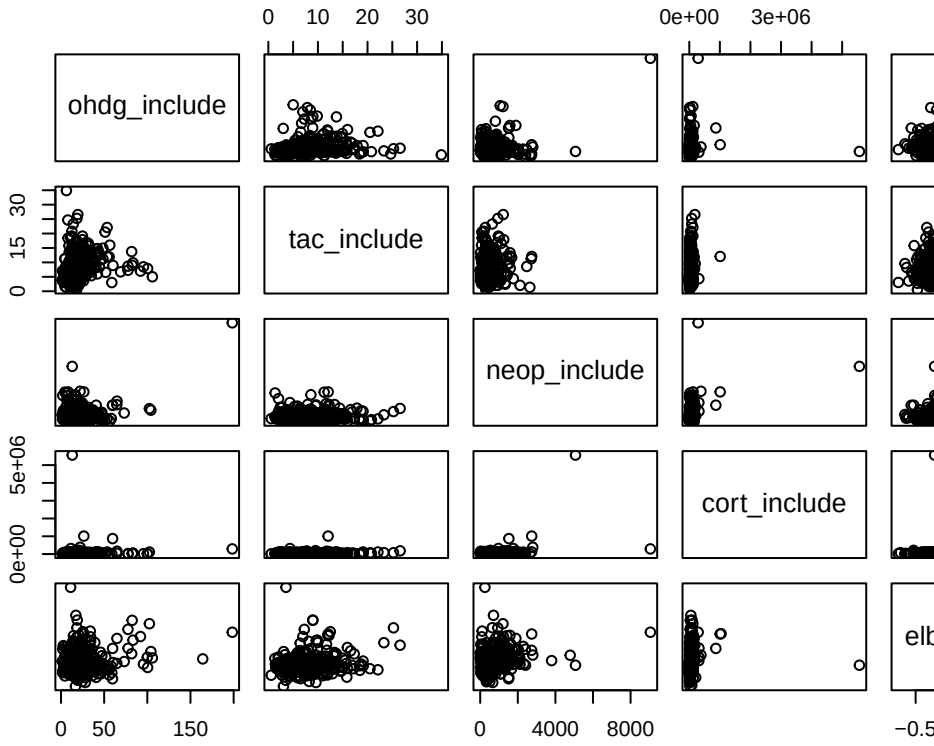
I created two markers for flagging each biomarker as an outlier. The first is based on -3:3 z-score. The second arranges the values of a biomarker in ascending order and calculates the distance between successive

points. My motivation was that the z-score (and IQR) method(s) seem very restrictive; several reasonable variables were cut out because the average of the biomarker was lower. The distance metric for 8-OHdG, for example, shows that the only data ponits that were very different from the values around it were the last two, which exhibited jumps of over 30 and 50, while all other distances were well below 5 or 10.

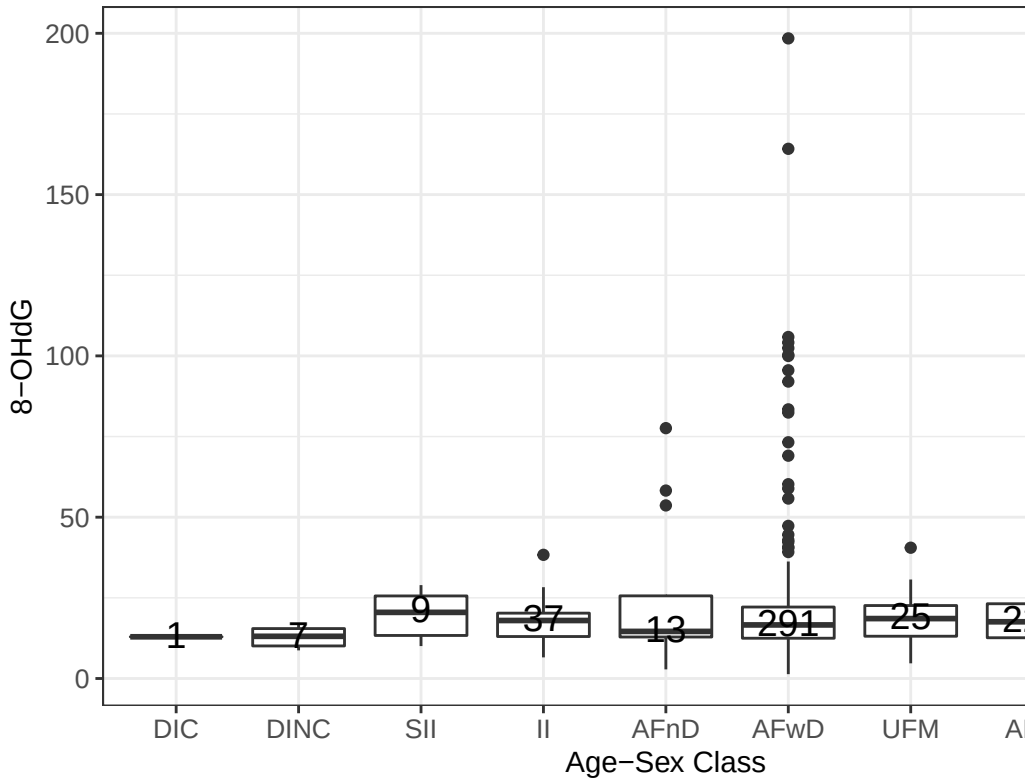


Based on these plots, I am going to flag the distance outliers as follows: * 8-OHdG: distances greater than 20 * TAC: distances greater than 1.5 * Neopterin: Distances greater than 500 * Cortisol: distance greater than 80,000 (800 on the plot)

Covariation

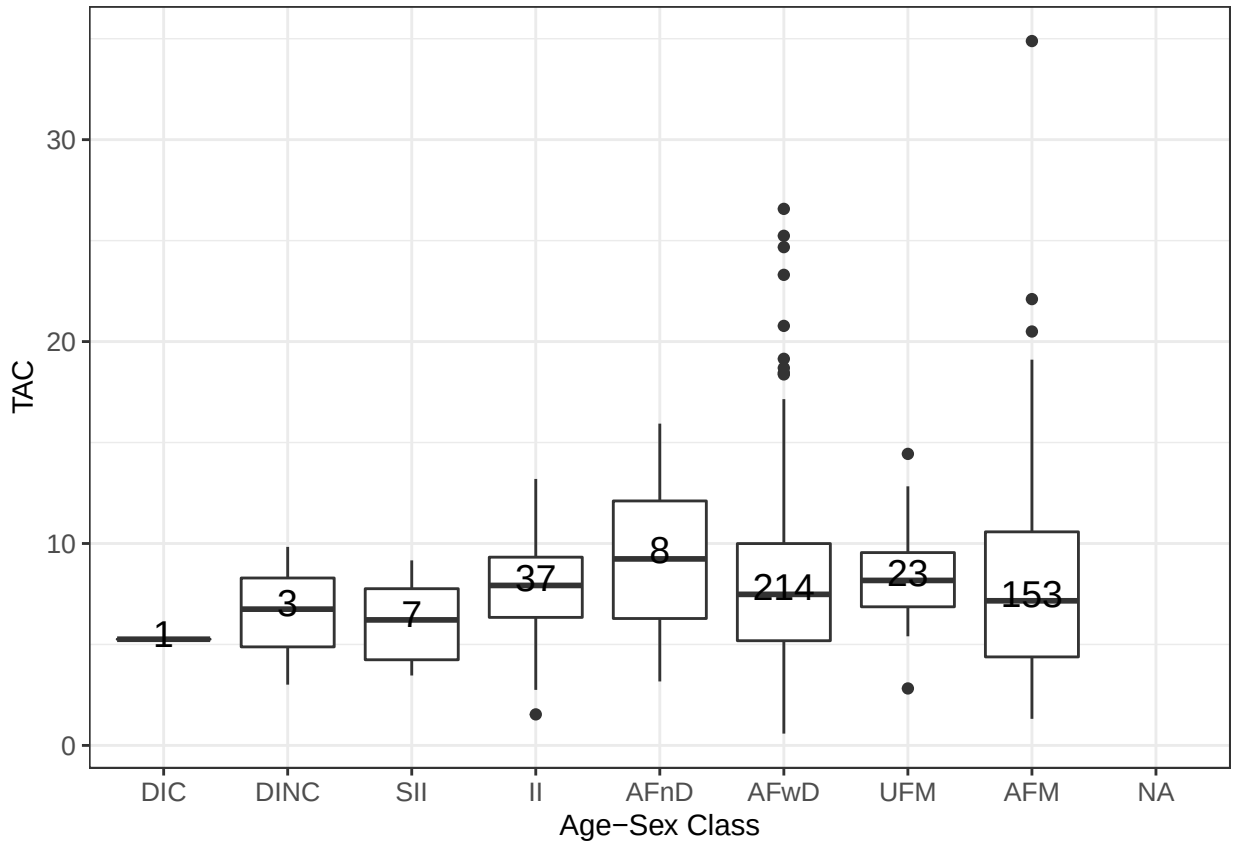


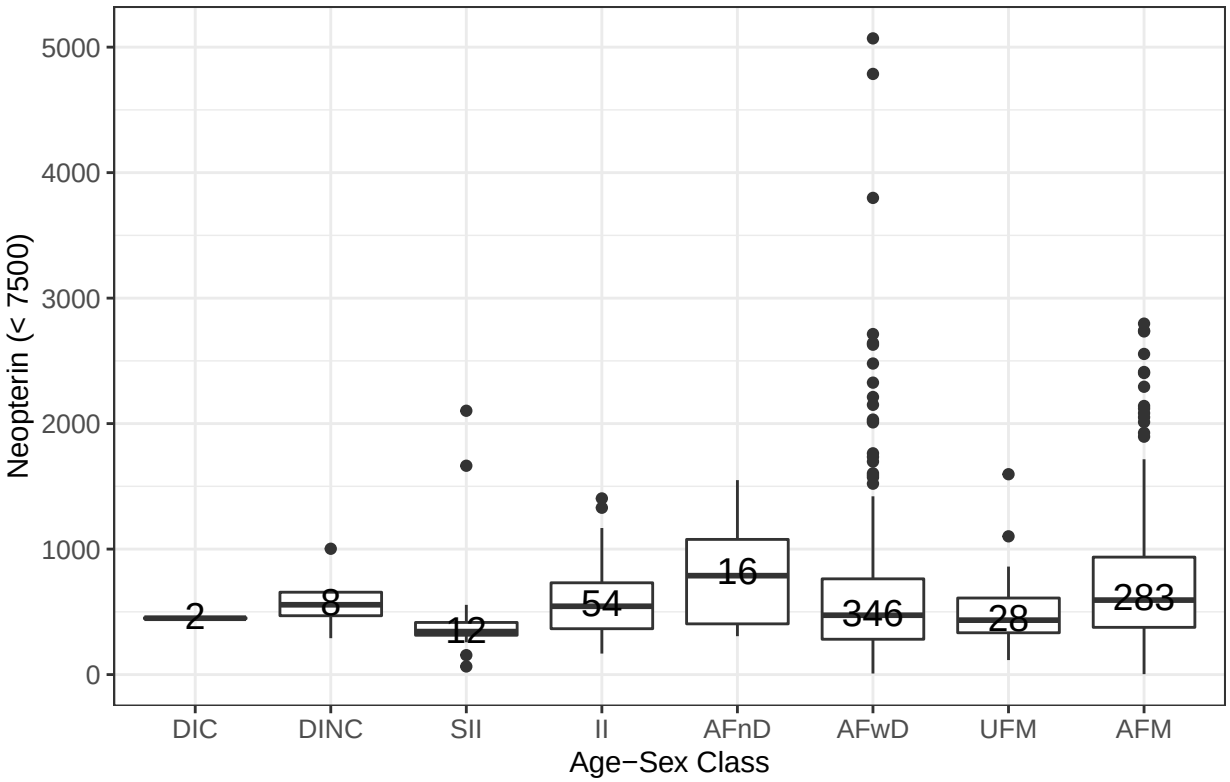
Covaration between Biomarkers



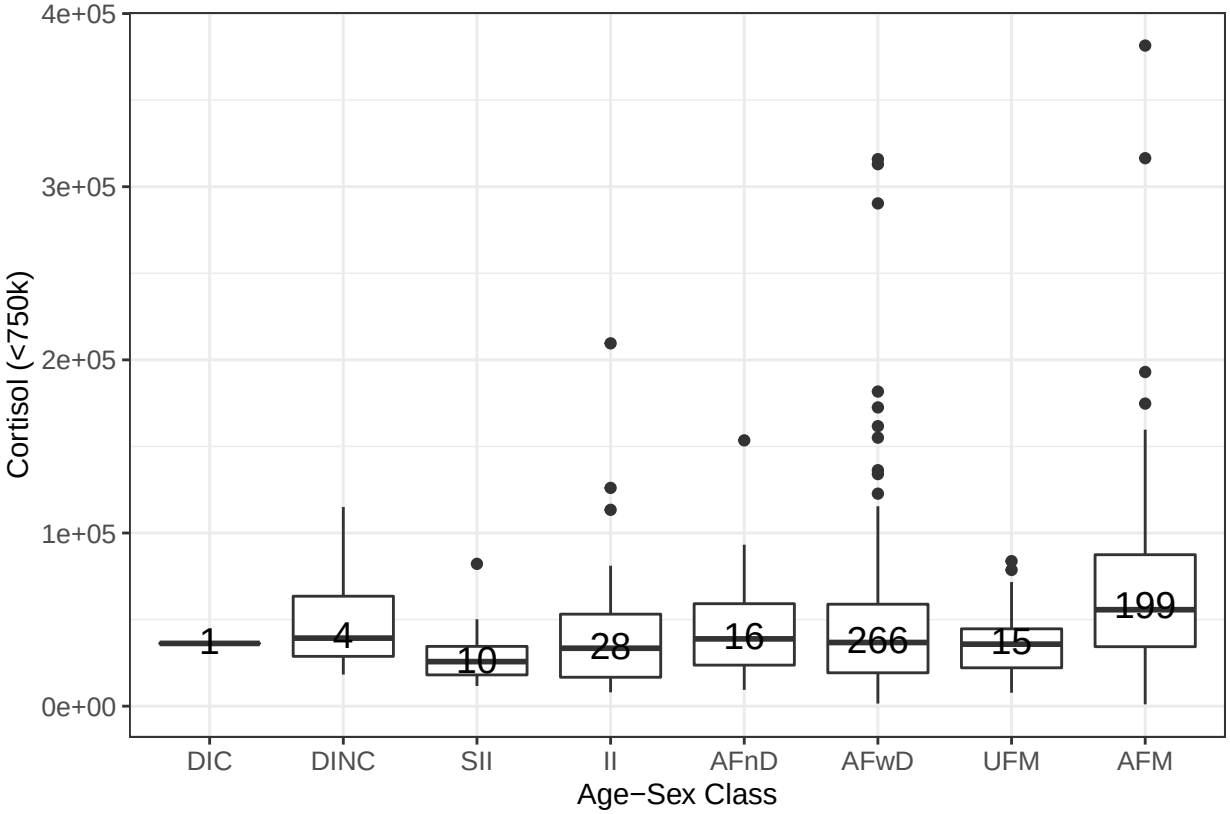
I cannot really identify a patt
chance for a lot of higher outli

Biomarkers by Age-Sex Class

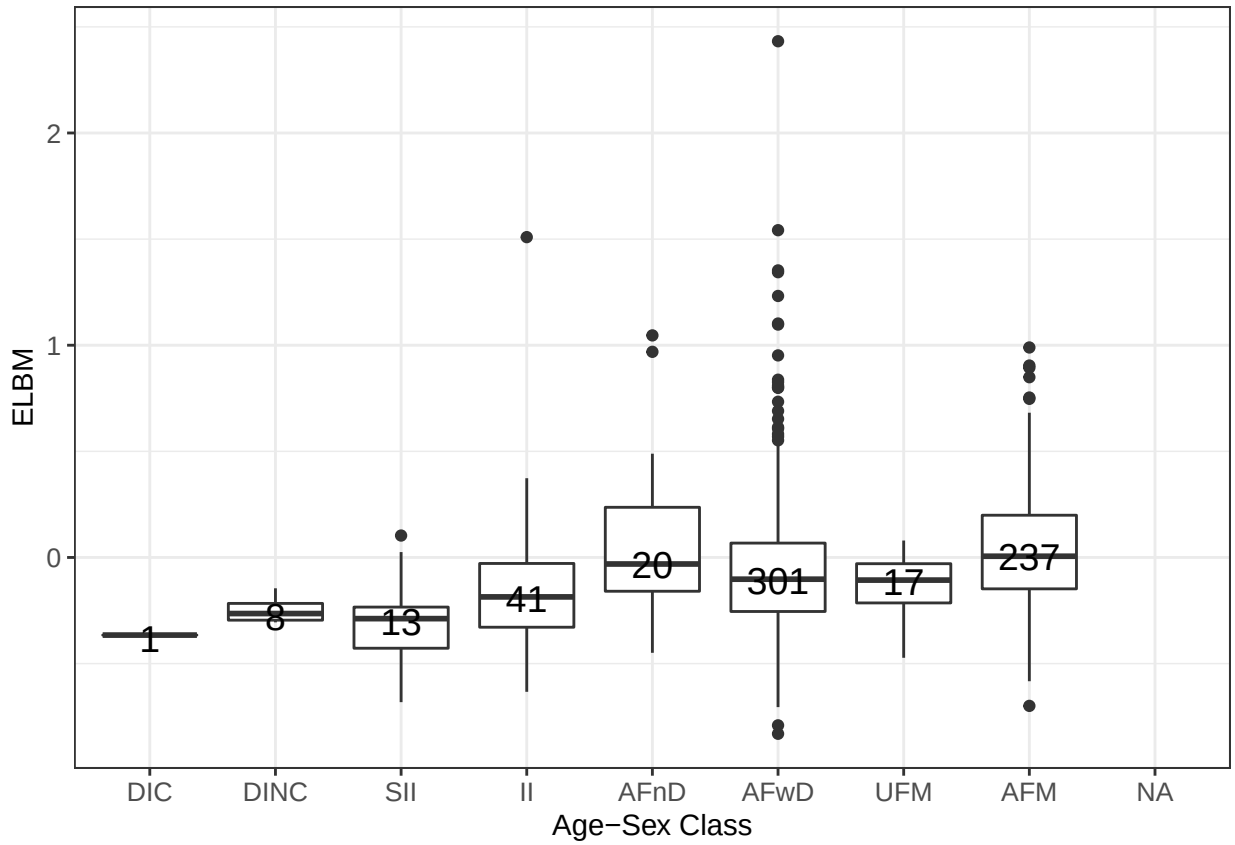




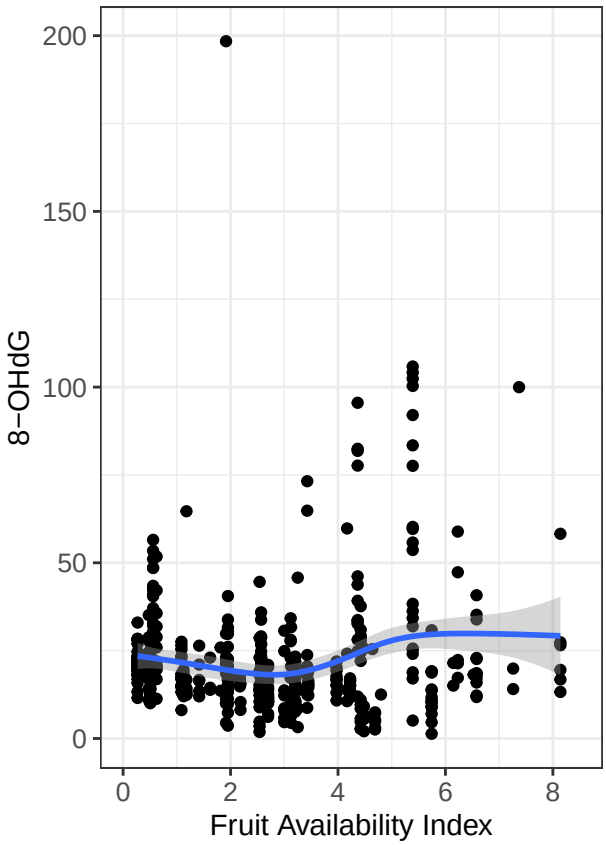
It does appear that adult flanged males higher median levels of neopterin

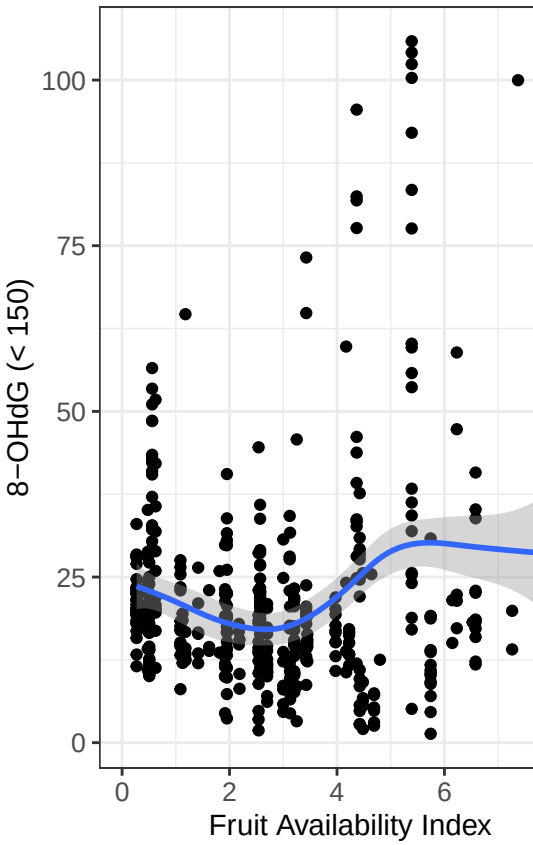


Flanged Males appear to have higher levels of cortisol

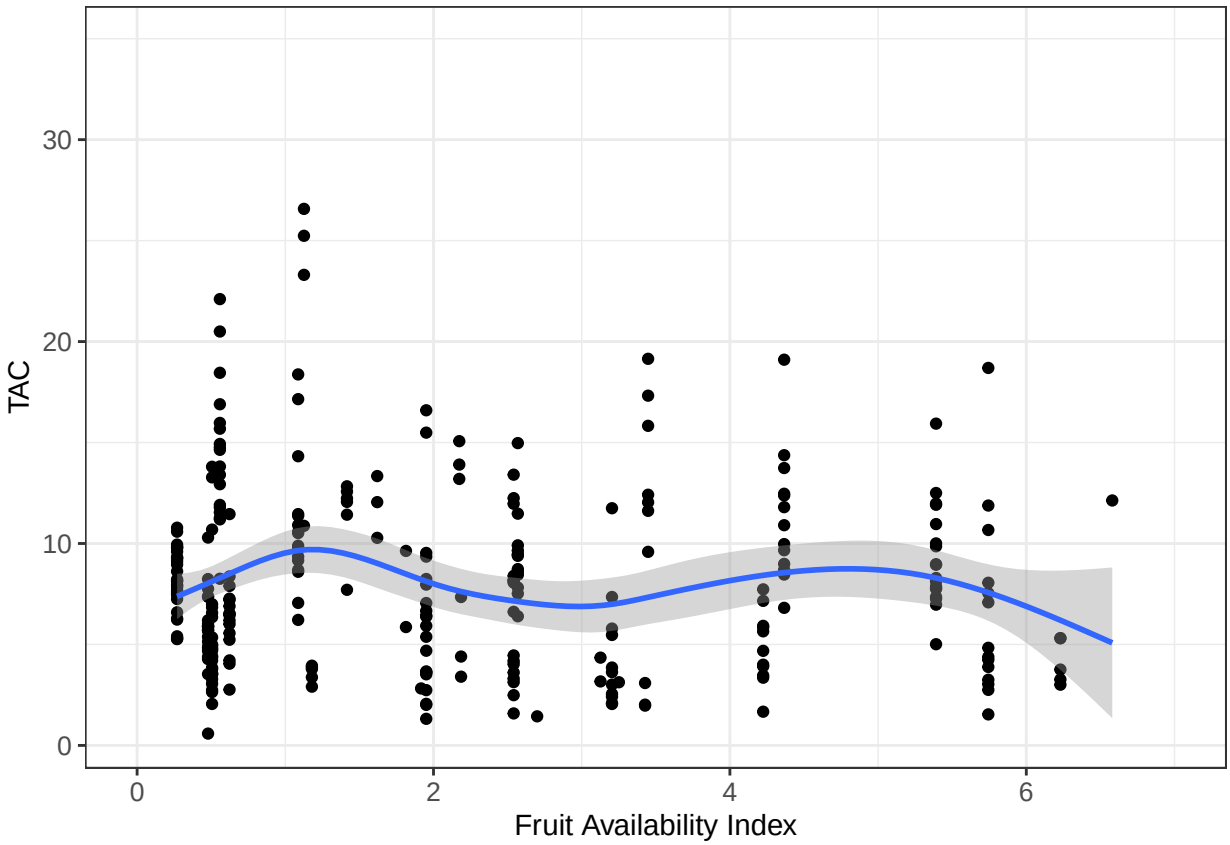


So it would appear that flanged males have higher cortisol and estimated lean body mass values, and potentially neopterin.

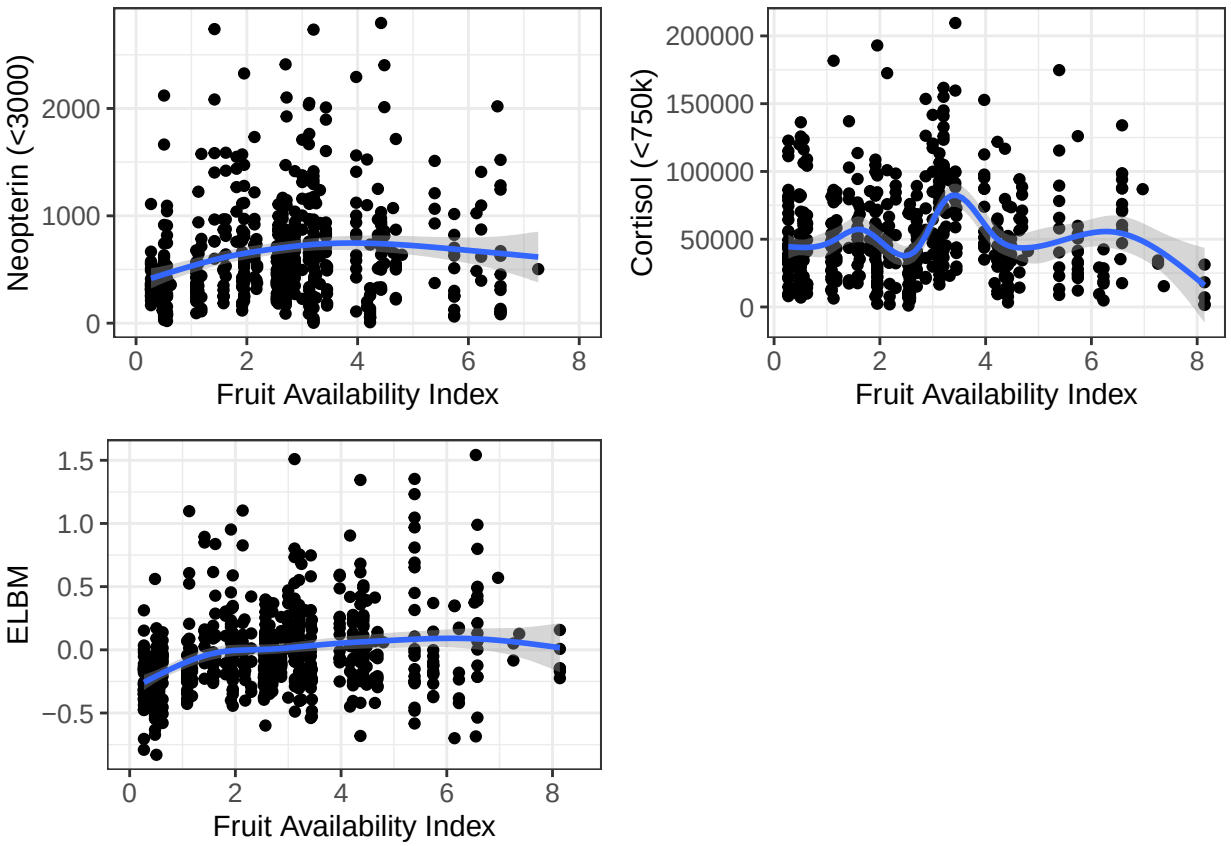




Biomarkers by FAI

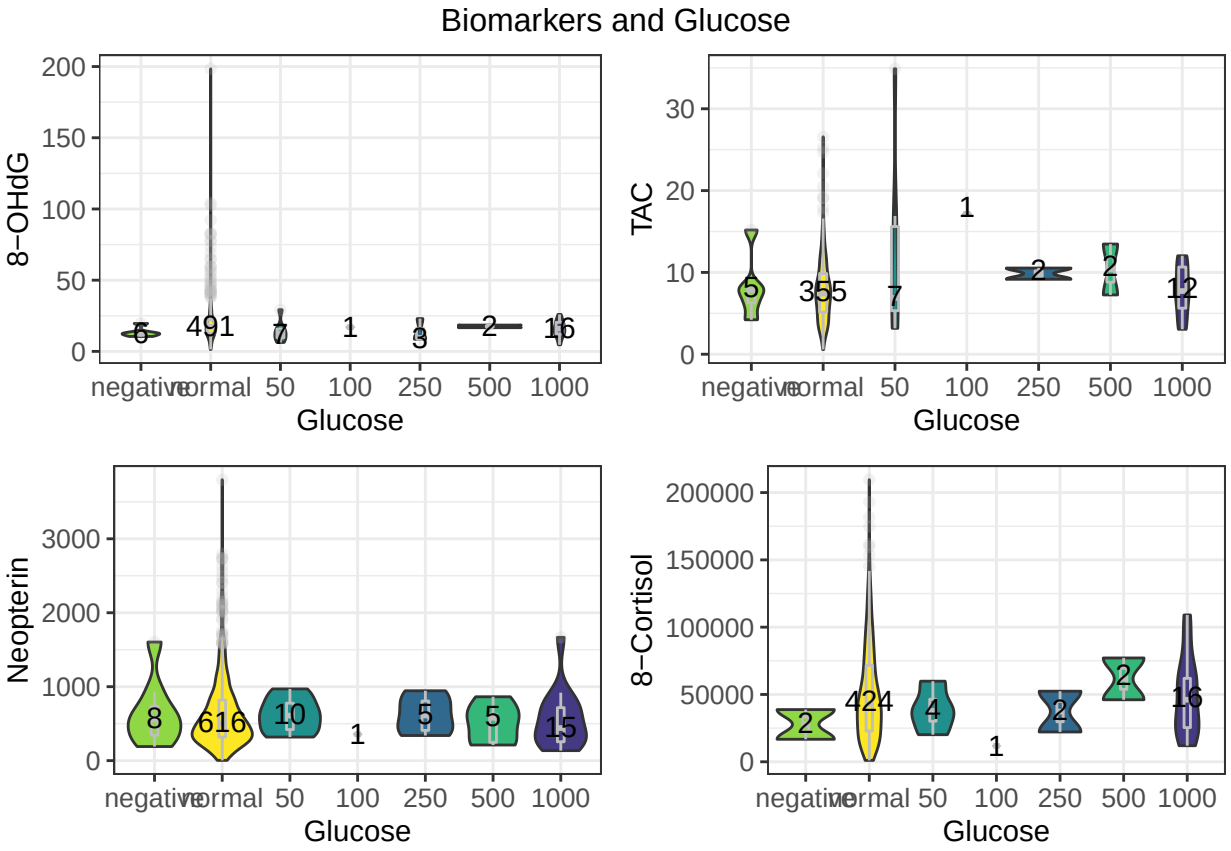


The increase in 8-OHdG related to FAI seems to be driven by a single point. Note that a much higher outlier exists elsewhere on the lower end of FAI so I doubt FAI is actually related to this at all. If I remove these two outliers, it would appear, ever so slightly, that there are two plateaus, where it starts out lower and then increases. However, this is not really suggestive of anything; FAI is not capturing information relevant to oxidative stress. TAC definitely does not seem to be associated with it. I mean maybe there is a bit of a bimodal peak. Maybme.

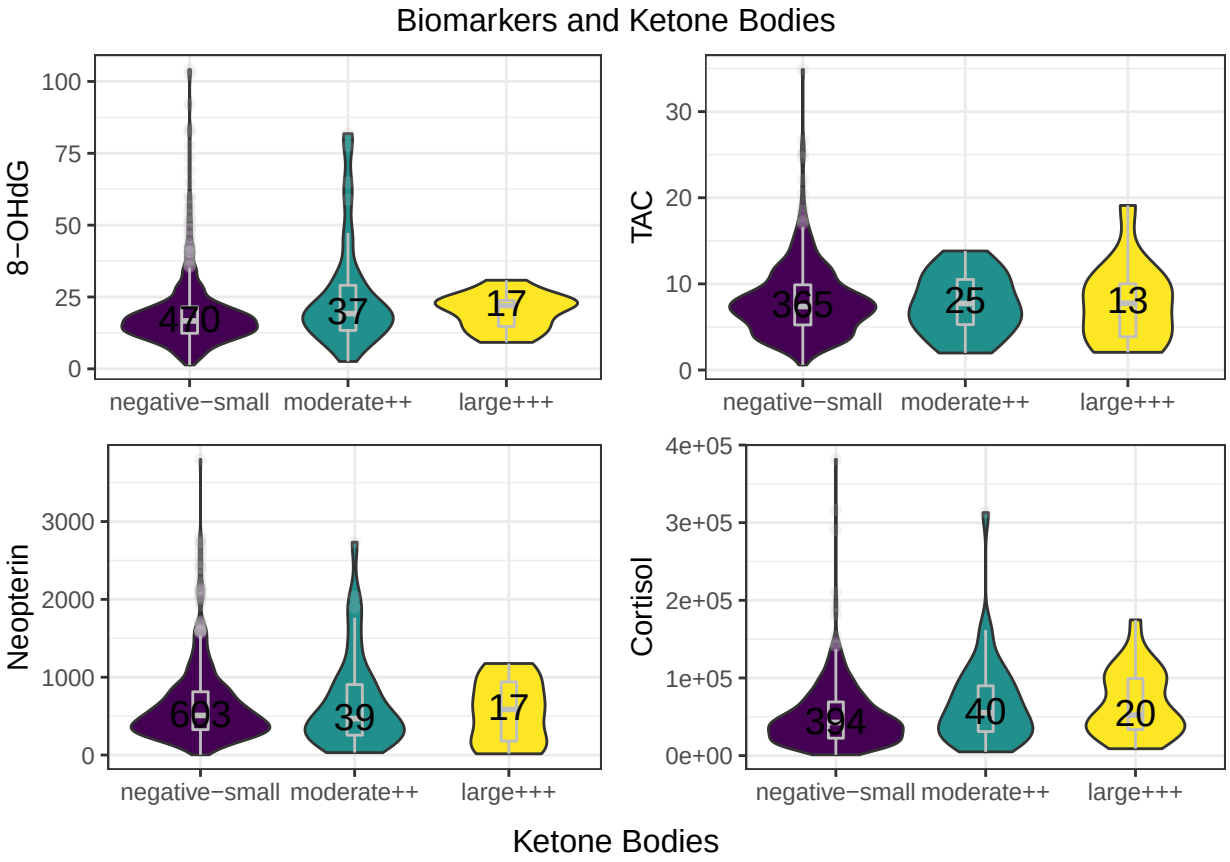


There were 4 outliers removed for neopterin. One was very extreme (above 7500). When I removed this and plotted, the current curve shape appeared. The removal of three more outliers showed this bow shape remain the same, if not strengthen a bit. Since the shape of association remained, I will stick with this visualization. For cortisol, there was one extreme outlier above 750k. 5 more observations were clustered between 250k and 750k, and I removed them from the plot. THe shape of relationship makes little sense regardless. I removed 1 outlier for ELBM that was above 2,but the shape remains the same. It was simply easier to view the panel plot this way.

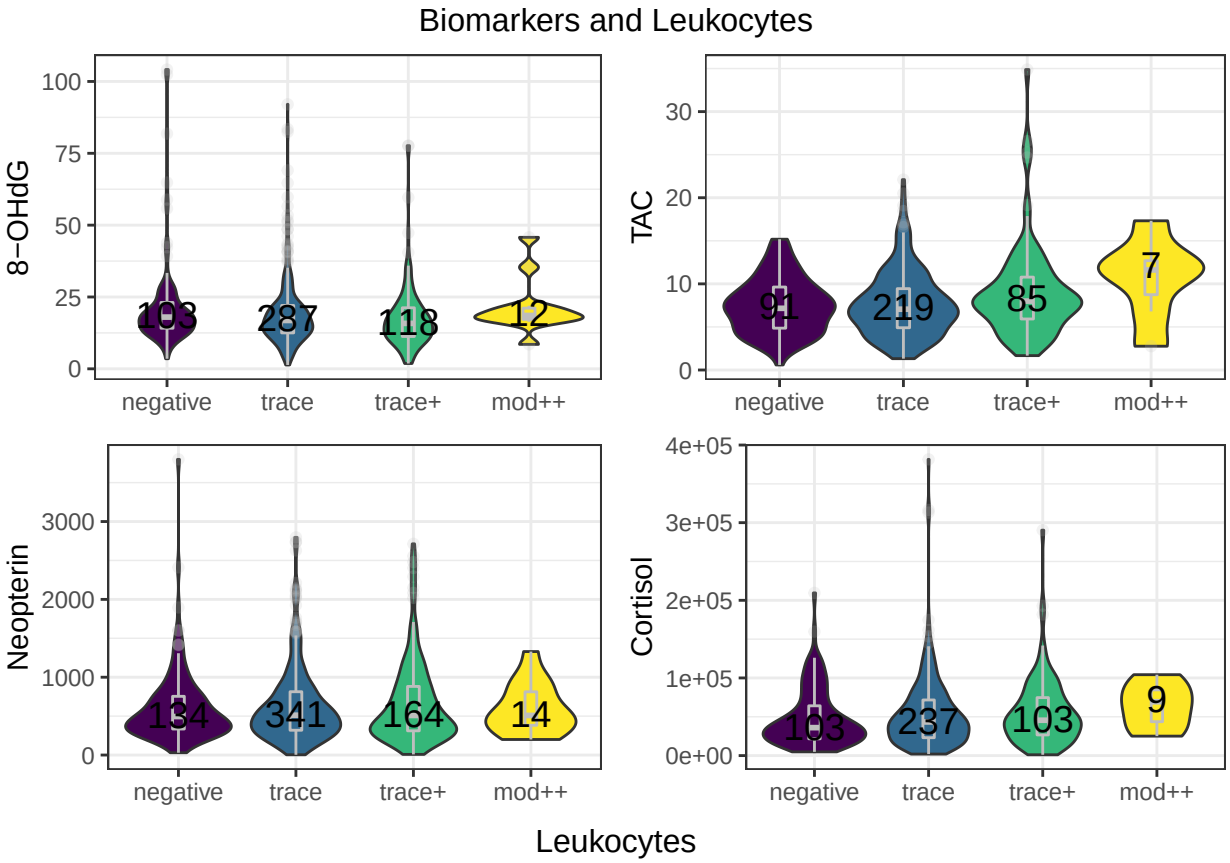
Biomarkers by Select Forest Chemstrip Data This part of the report was inspired by a paper talking about the effects of high glucose induced increases in oxidative stress episodes on myogenesis, particularly acceleration (Liu et al. 2020).



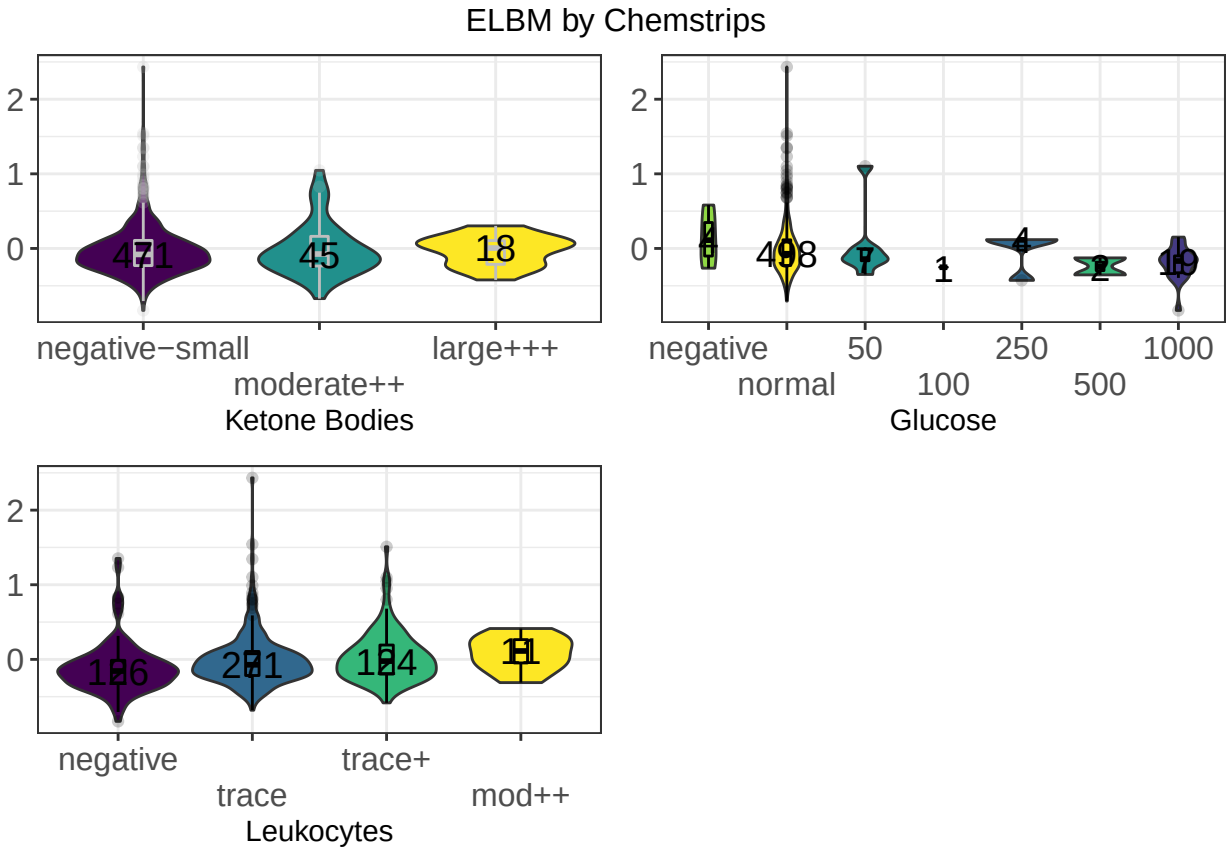
Given the tiny sample sizes with a glucose reading above normal, I doubt these distributions have any true relevance.



I'm not sure what to say about ketones. There are just. . . some samples but there is such significant class imbalance that i'm not sure I could tease apart whether ketones were affecting these biomarkers.



I'm not particularly seeing a relationship herre. There looks like a steady positive increase in TAC might occur due to circulating leukocytes. The highest category always seems to be higher, but it has a full order of magnitude smaller a sample size so extreme caution should be taken in drawing any conclusions.



Very few samples are under the moderate and large categories of ketone bodies. It would appear that some degree of increase is found in the large category. I can't really say anything about glucose. Seeing as all categories fall within the same rough range, particularly of the negative category. If these data are accurate, I'd say there really isn't an effect. Likely the strips are simply not effective at measuring glucose. When it comes to white blood cells, there is some evidence of an increase in lean body mass. This begs the question of what the relationship between neopterin will be.

Additional Supplementary Information

How many samples were collected by each field worker? This first table will include all assayed samples, broken down by type of assay. All samples are included regardless of CV, whether it was a first morning void, whether it is too dilute, or taking any other considerations.

Table 5: Table 1b: All assayed samples by which field worker collected them

biomarker	Who_collected_clean	num_collected	prop_collected
8-OHdG	Abuk	182	24.53
	Abuk, Alysse	1	0.13
	Allie	7	0.94
	Alysse	11	1.48
	Anna	2	0.27
	Beth	1	0.13
	Caco	2	0.27
	Cecilia	2	0.27
	Conor	3	0.40
	Conor, Ellie	1	0.13
	Daniel	4	0.54
	Ellie	6	0.81
	Erin	1	0.13
	Idun	135	18.19
	Idun and Tim	1	0.13
	Idun/Alysse	1	0.13
	Isman	82	11.05
	Jade	2	0.27
	Jade, Hanon?	1	0.13
	Julia	13	1.75
	Kumpo	1	0.13
	Lady	1	0.13
	Manon	2	0.27
	Misdi	1	0.13
	Paige	7	0.94
	Pawel, Kumpo	1	0.13
	Rahmadt	32	4.31
	Rahmatol	1	0.13
	Rebecca	10	1.35
	Rumaan	1	0.13
	Shaylyn	9	1.21
	Sonja	1	0.13
	Suwi	102	13.75
	Suwi, Tim	1	0.13
	Suwi/Isman	1	0.13
	Tim	4	0.54
	Tono	80	10.78
	Tono, Wendy	1	0.13
	Unknown	1	0.13
	Wendy	20	2.70
	Wilhelm	3	0.40
	Will	1	0.13
	Yann	3	0.40
	Abuk	207	20.80
	Abuk and Tim	1	0.10
	Abuk, Alysse	1	0.10
	Allie	13	1.31
	Alysse	78	7.84
	Alysse/Idun	9	0.90
	Anna	7	0.70
	Awan	1	0.10
	Beth	4	0.40
	Bonnie	1	0.10
	Brigitte	1	0.10
	Bunga	2	0.20
	Caco	17	1.71
	Cecilia	4	0.40