RHICCA Baseline Analysis

10 August 2020

RHICCA Baseline Data: Visual Summary

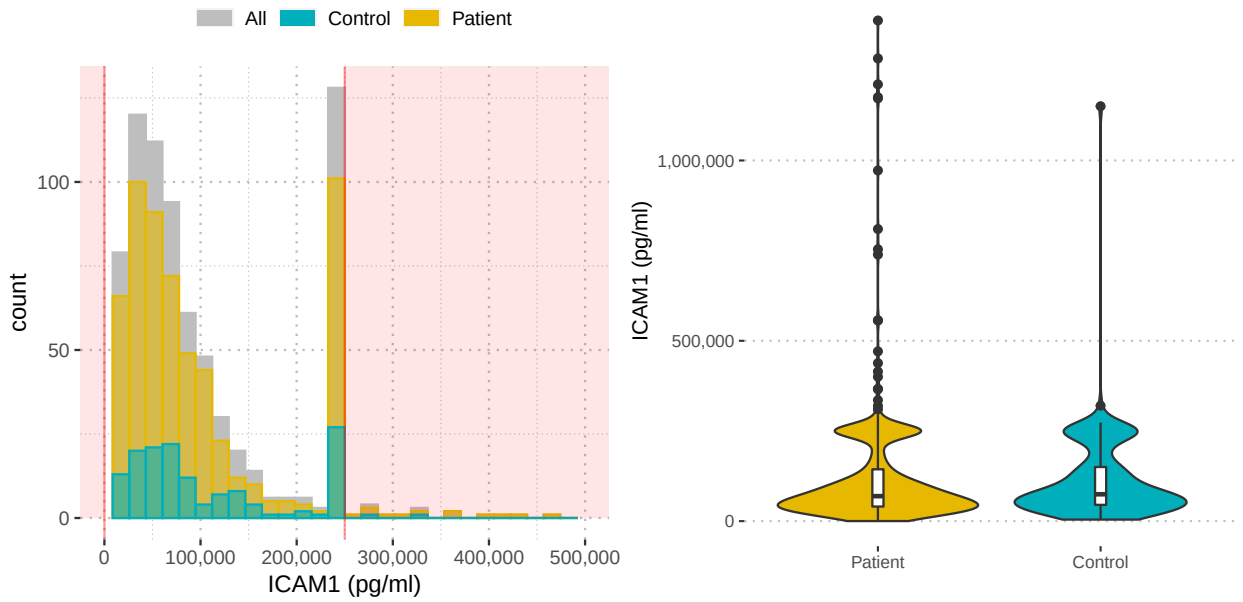
This document contains graphs and charts showing the RHICCA data collected at baseline, including each of the four biomarkers and how they may vary depending on ART status, gender, and age. Of the 827 samples in the baseline data, 766 could be successfully matched with clinical data, though some errors and outliers may still be present.

The data for each biomarker is either left-censored, right-censored, or both, due to the detection limits imposed by the Luminex Assay. Regions outside of these limits are shaded in red in the histograms, although there is still some viable data within these regions.

ICAM1: ART vs Control

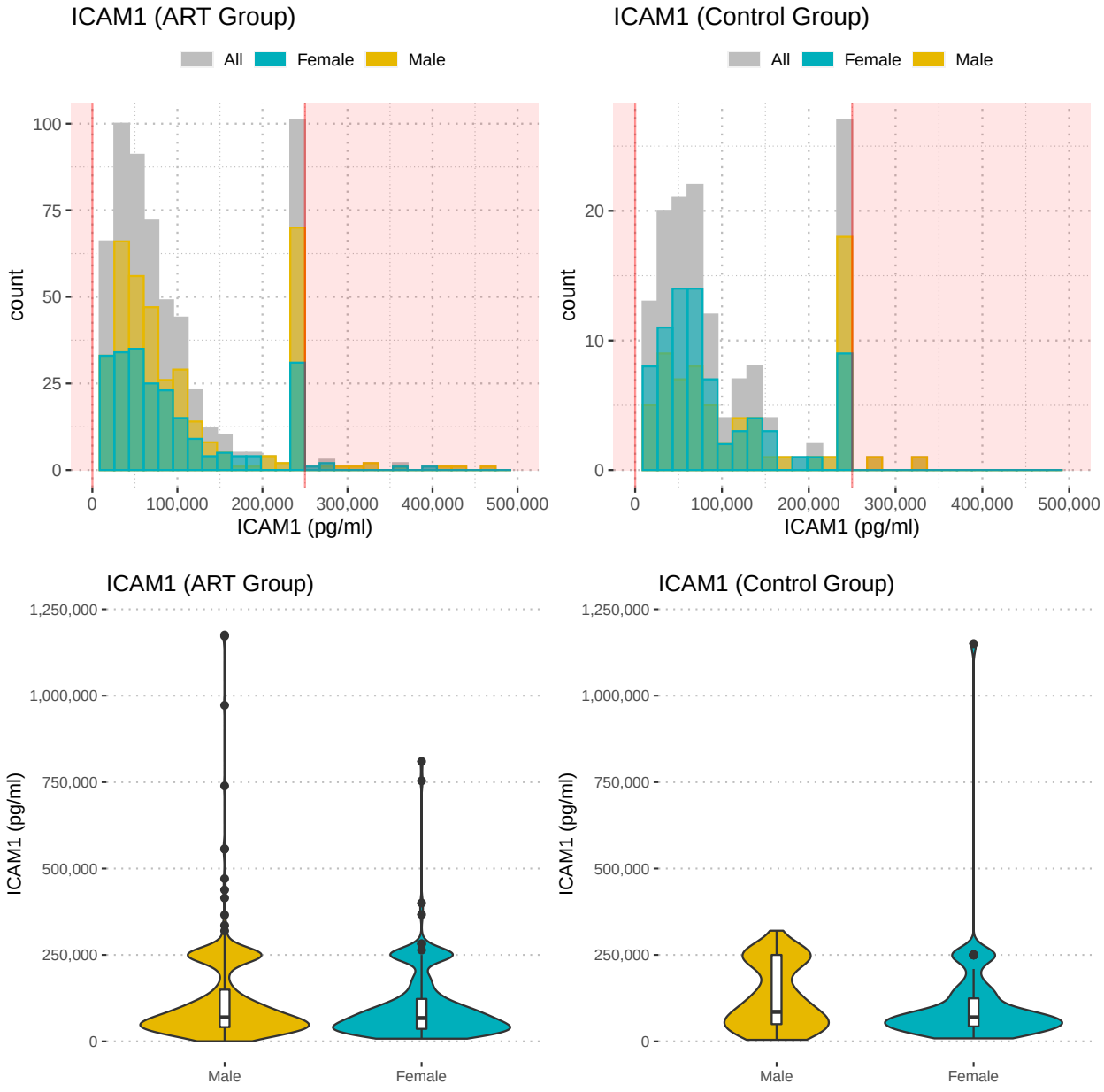
Some extreme values have been left off of the histograms below, as the significant skew of the data makes an informative histogram more difficult to construct. The violin plots, however, contain all values of the data. We see that overall, the shapes of the ICAM1 distributions for ART and Control subjects look similar, although most extreme values belong to the ART patients.

The upper detection limit of 250 000 results in a vertical bar on the x-axis where values listed as ">250 000" were imputed to that limit



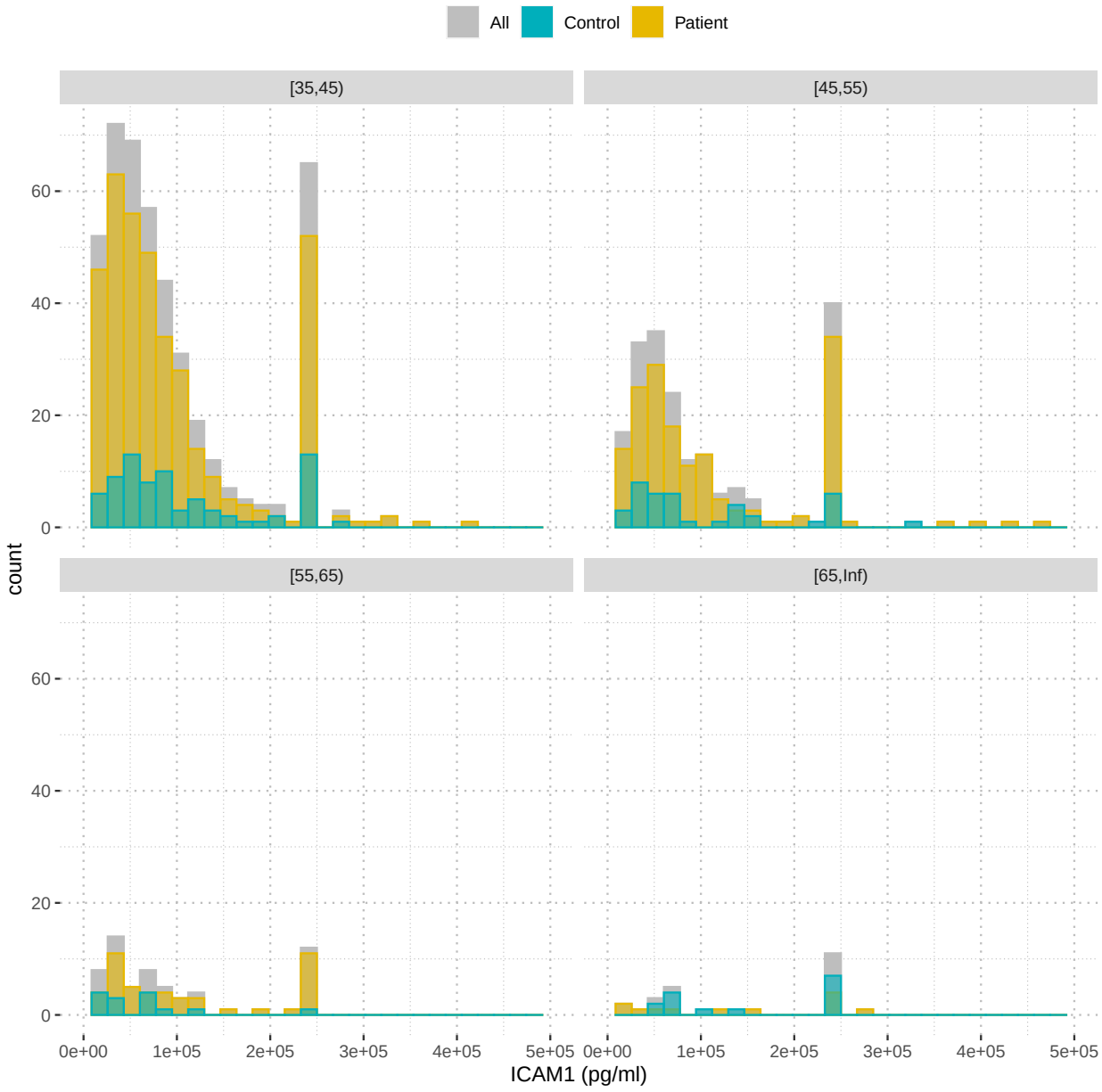
ICAM1: Sex

The data, when separated by ART status and gender, do not seem to show significant differences in distribution, although male ART patients do seem to experience a somewhat greater increase in ICAM1 levels compared to control subjects than women do.



ICAM1: Age

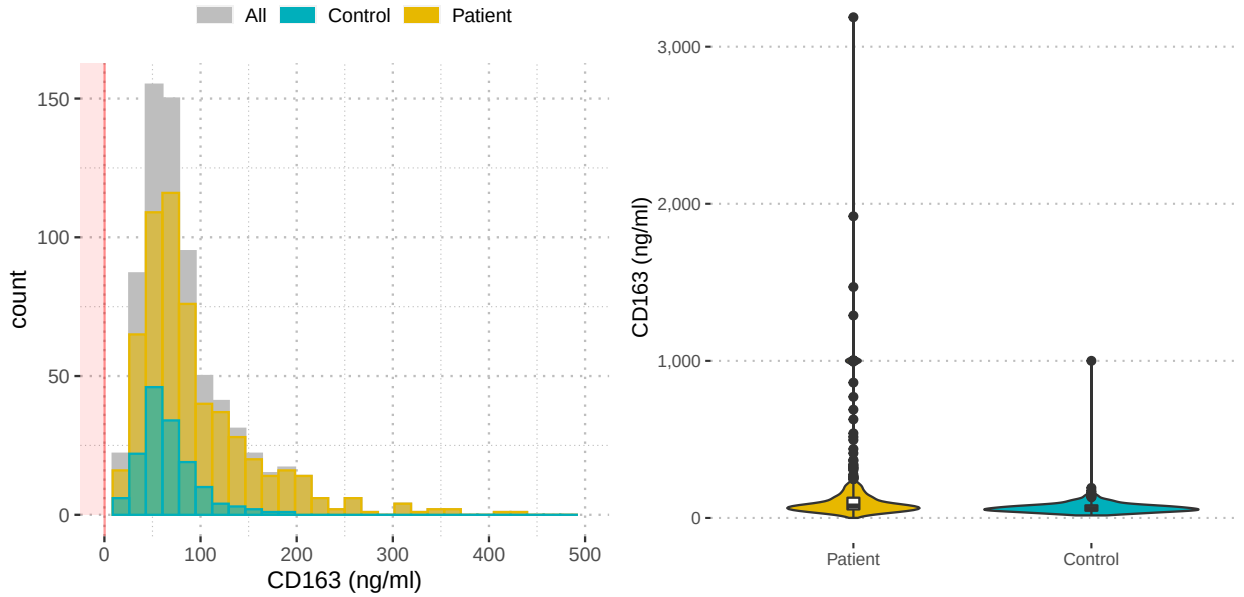
Separating by age group does not seem to show significant differences in distribution



CD163: ART vs Control

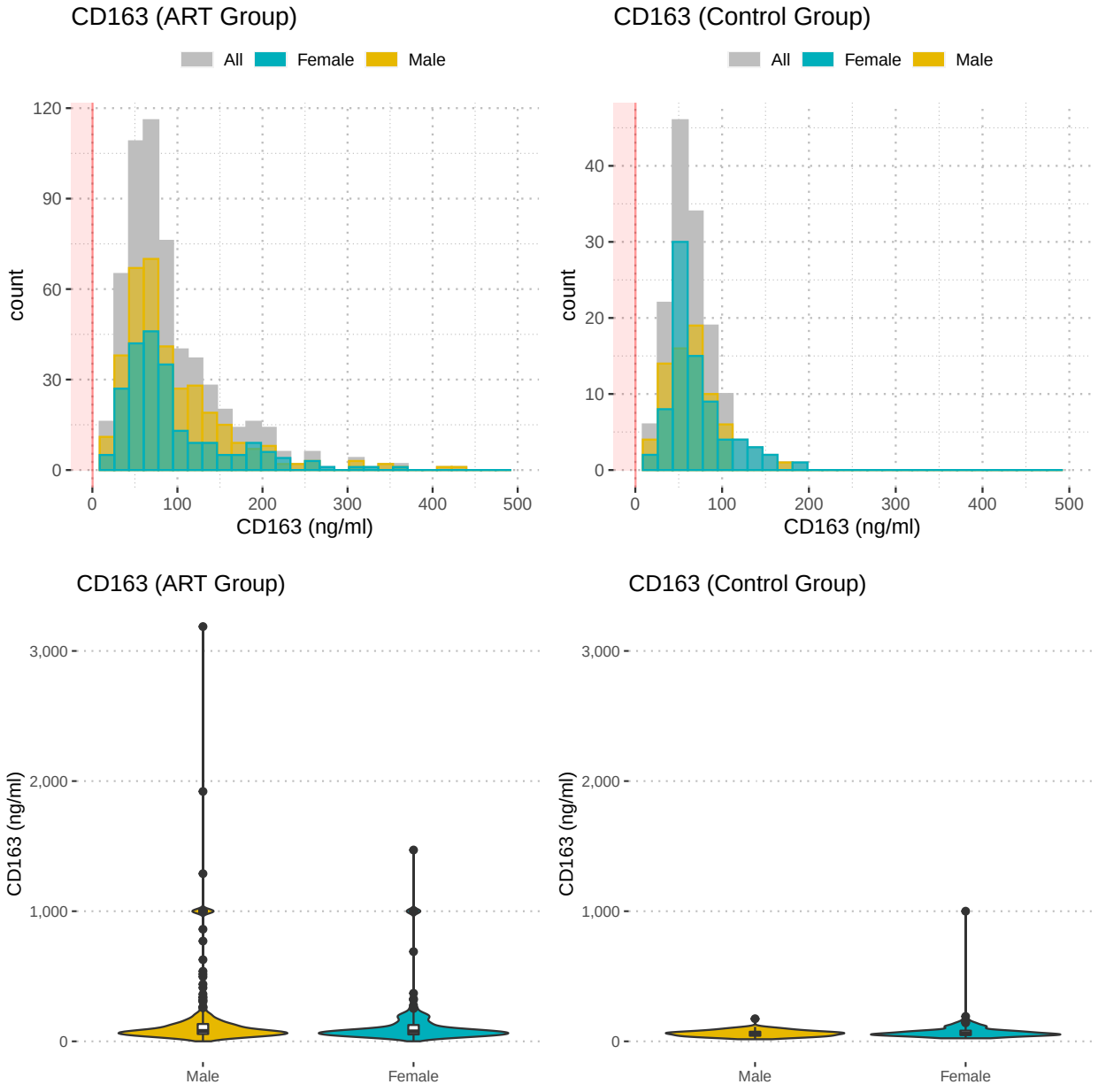
Once again, some extreme values have been excluded from the histogram. We see perhaps a slight tendency toward higher CD163 values in the ART patients but no strong evidence of a difference in distribution between the ART and control subjects.

Only a handful of values lie outside the detection limits.



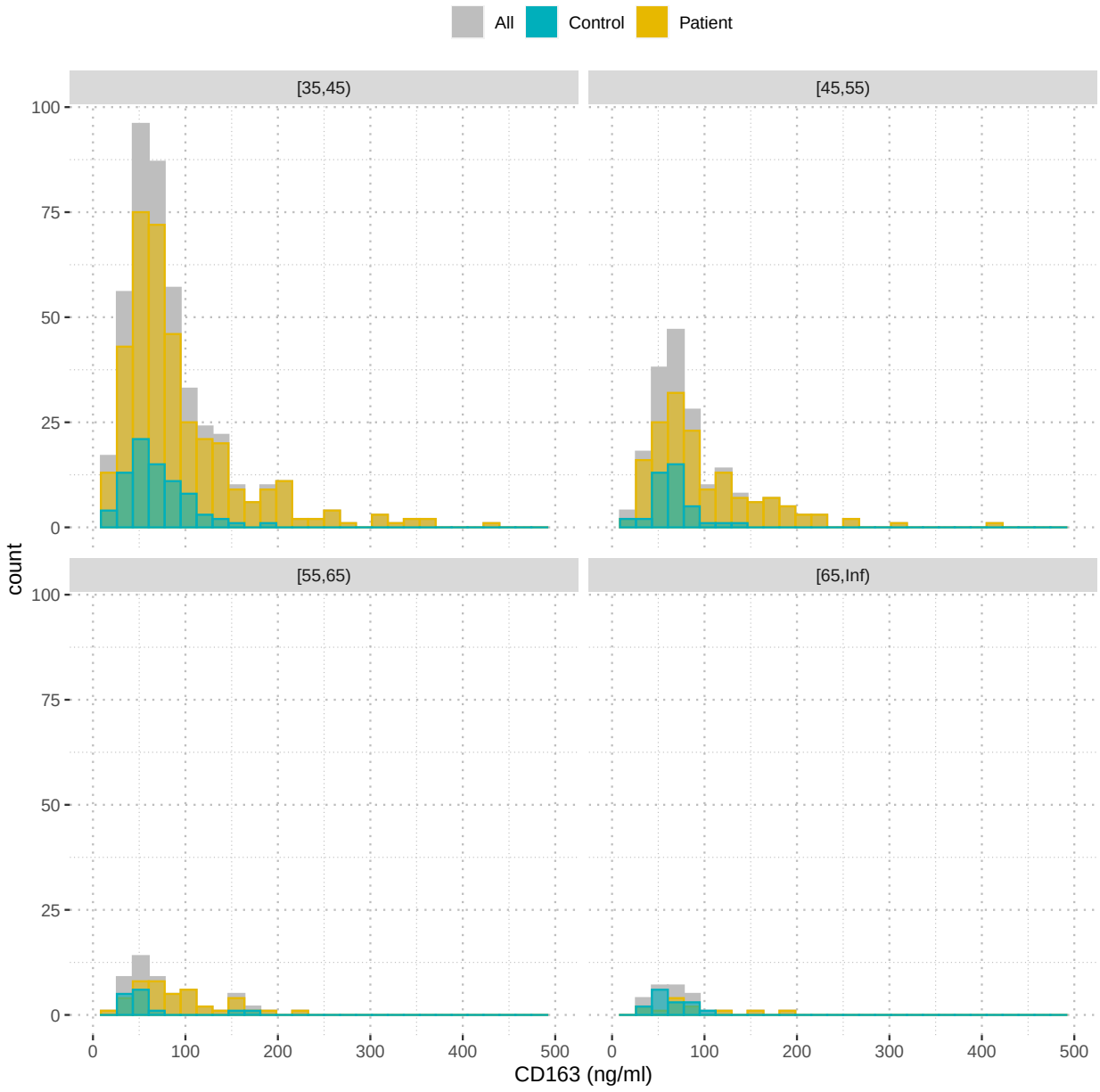
CD163: Sex

The data, when separated by ART status and gender, do not seem to show significant differences in distribution, although once again we see that male subjects are responsible for most of the recorded increase in CD163 levels that ART patients experienced.



CD163: Age

There does not appear to be a difference in distributions between age groups

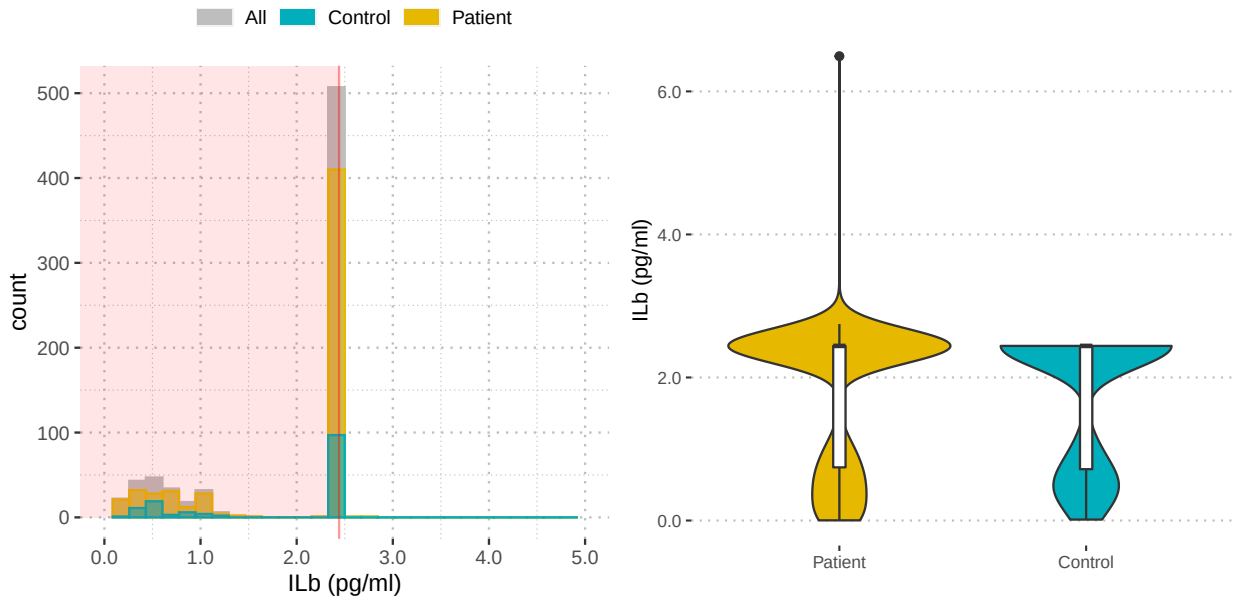


ILb: ART vs Control

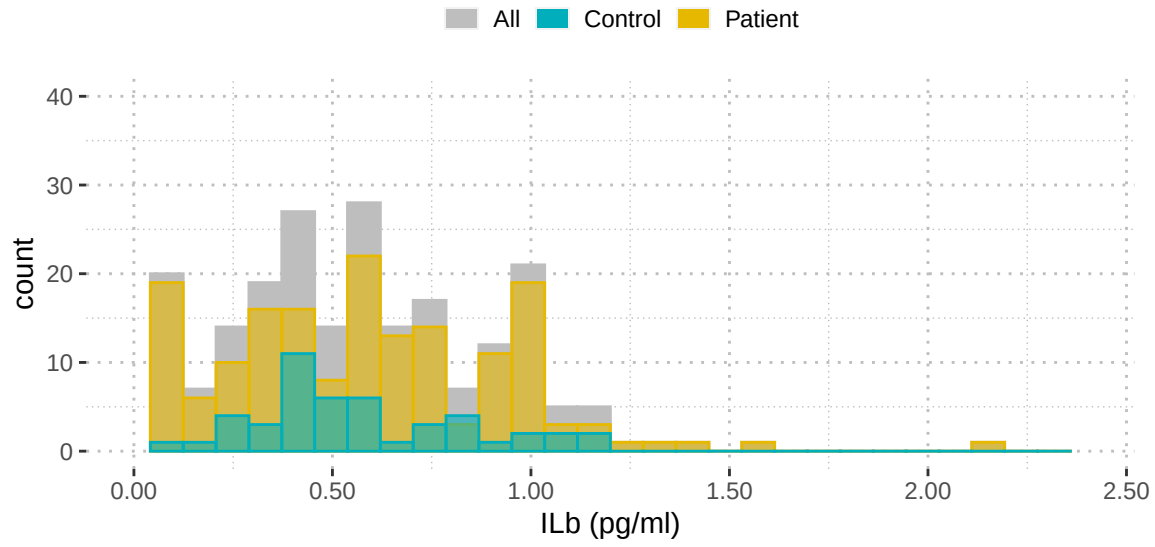
Two extreme values have been excluded from the histogram, including one very high value of 131828.8 pg/ml that has also been excluded from the violin plot. Waiting on information from Dumisani and Dumizulu about the validity of this measurement but as it is identical to the ICAM1 value from the same row and the rest of the ILb measurements fall between 0 and 7 pg/ml, it seemed likely to be an error and thus counterproductive to include it in any plots for the time being.

We see that the charts are dominated by a large peak at x=2.44140625, the lower detection limit for this biomarker, as a result of the fact that 507 out of 766 values (about 2/3 of the data) were given as "<2.44140625". The vast majority of the remaining values actually fall below the lower detection limit as well, although these can be presumed to still be valid.

The relatively small number of non-censored observations makes discerning patterns in the distributions difficult, but the distributions for ART and control patients do seem similar.

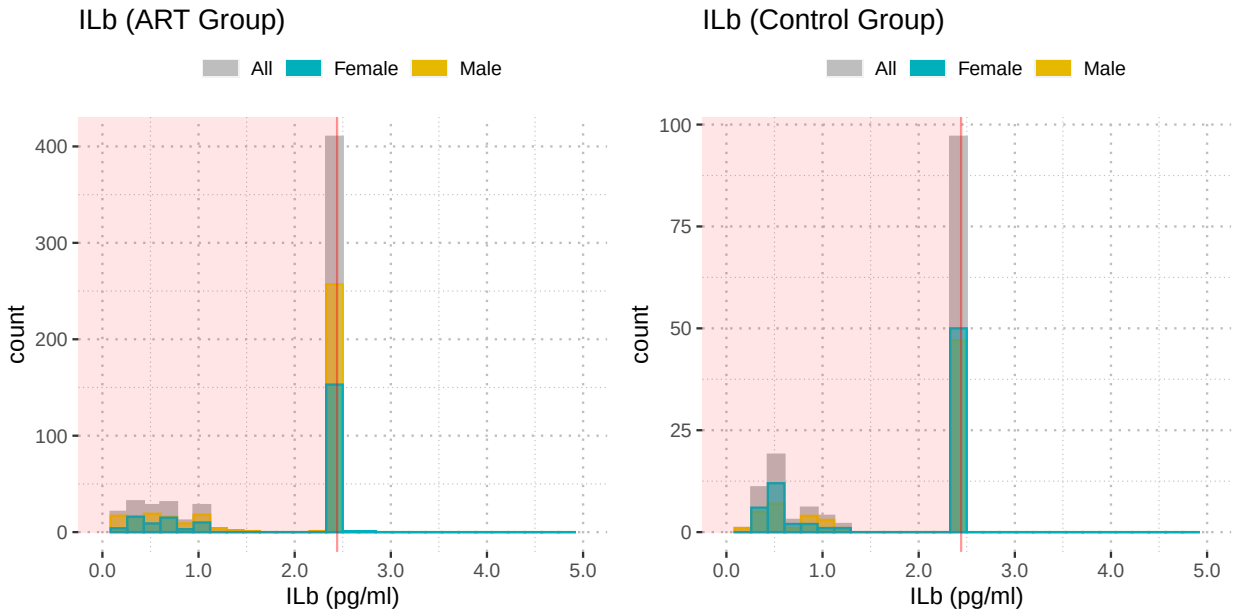


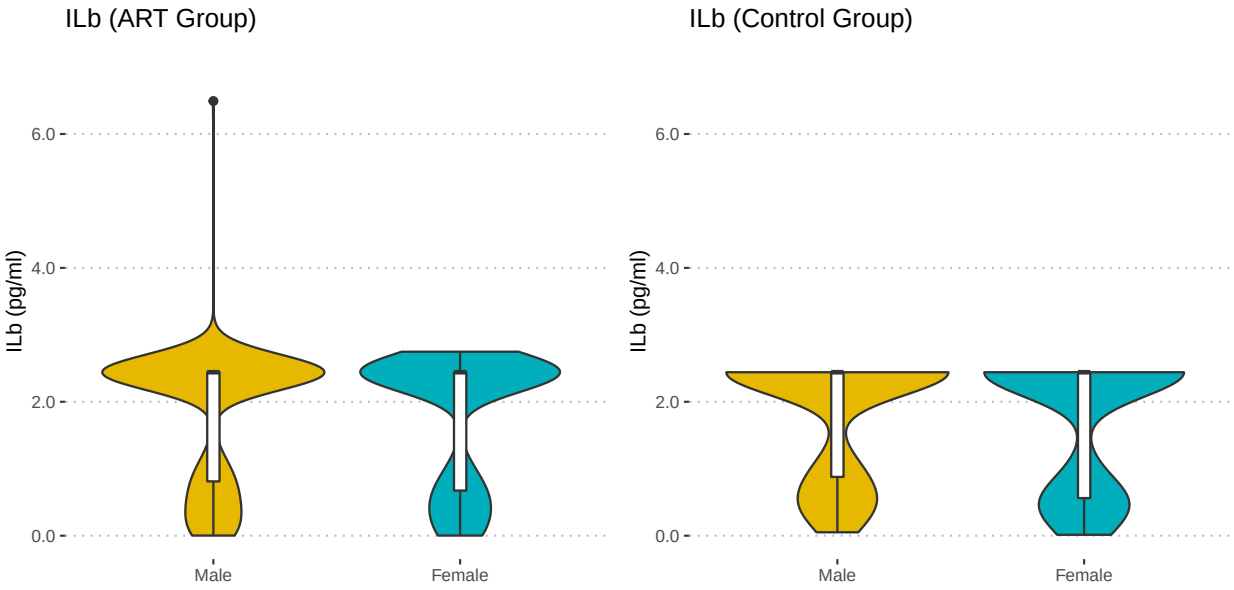
a closer look:



ILb: Sex

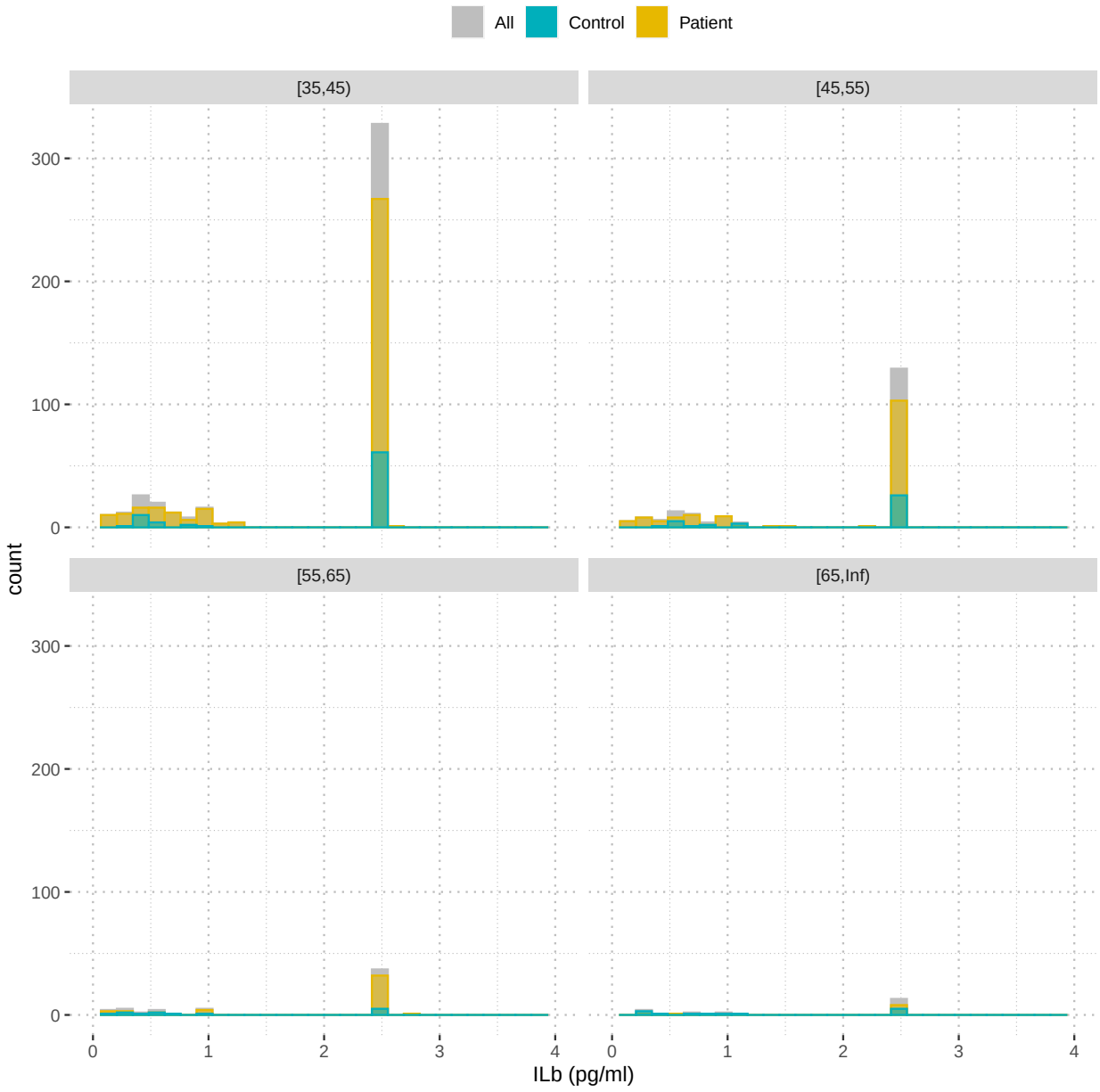
The data, when separated by ART status and gender, do not seem to show significant differences in distribution.





ILb: Age

Separating by age group does not seem to show significant differences in distributions

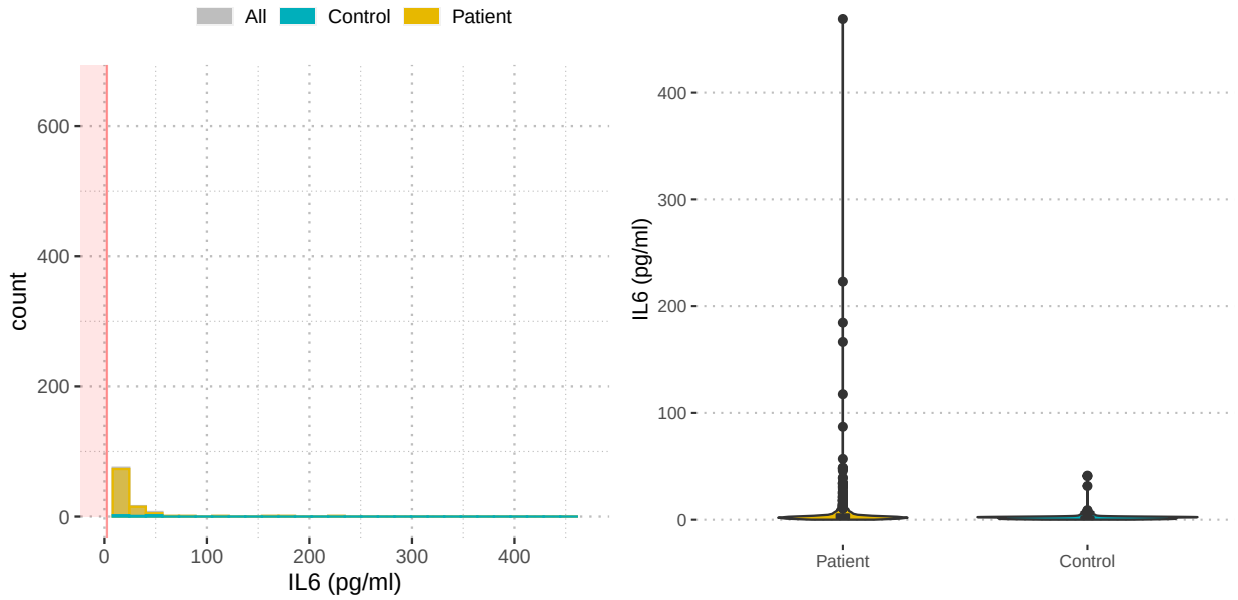


IL6: ART vs Control

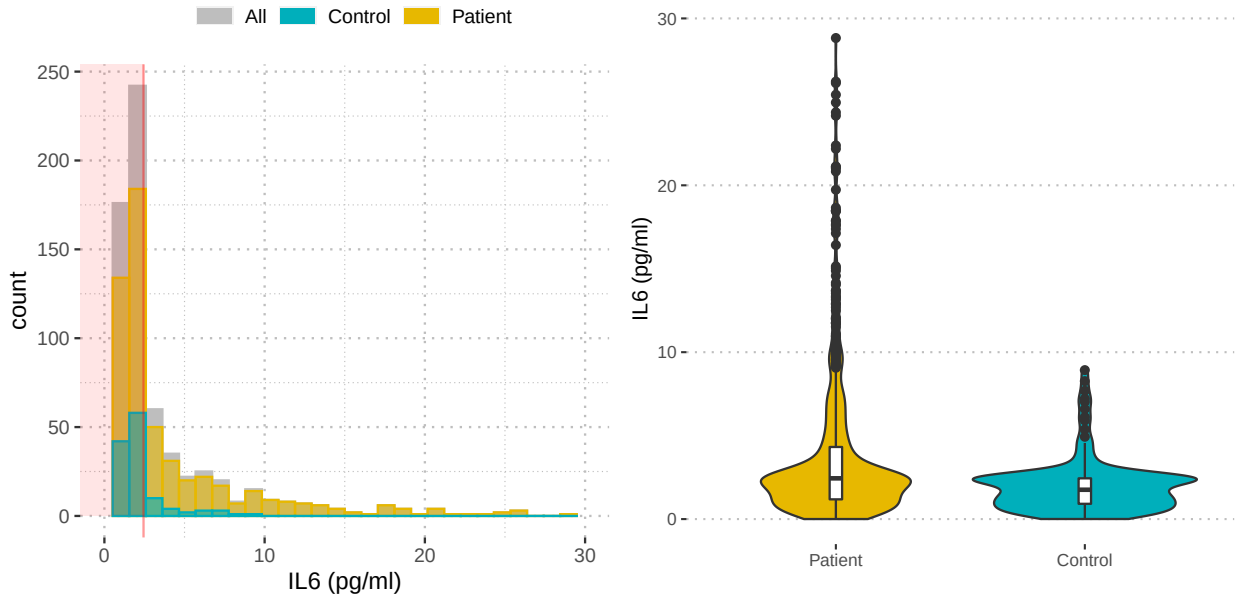
Once again, a number of extreme values have been removed from the histogram. Most of the data lies below 30 pg/ml, with the third quartile value being 4.0689, although the highest values are in the hundreds. This makes a complete violin plot difficult to interpret, but these extreme values do not seem to be errors, but rather high values from a right-skewed distribution. Therefore, a truncated and non-truncated version of both the histogram and violin plots are shown below

Although both distributions are right-skewed, we see distinctly higher values for ART patients than we do for control subjects.

Non-truncated:

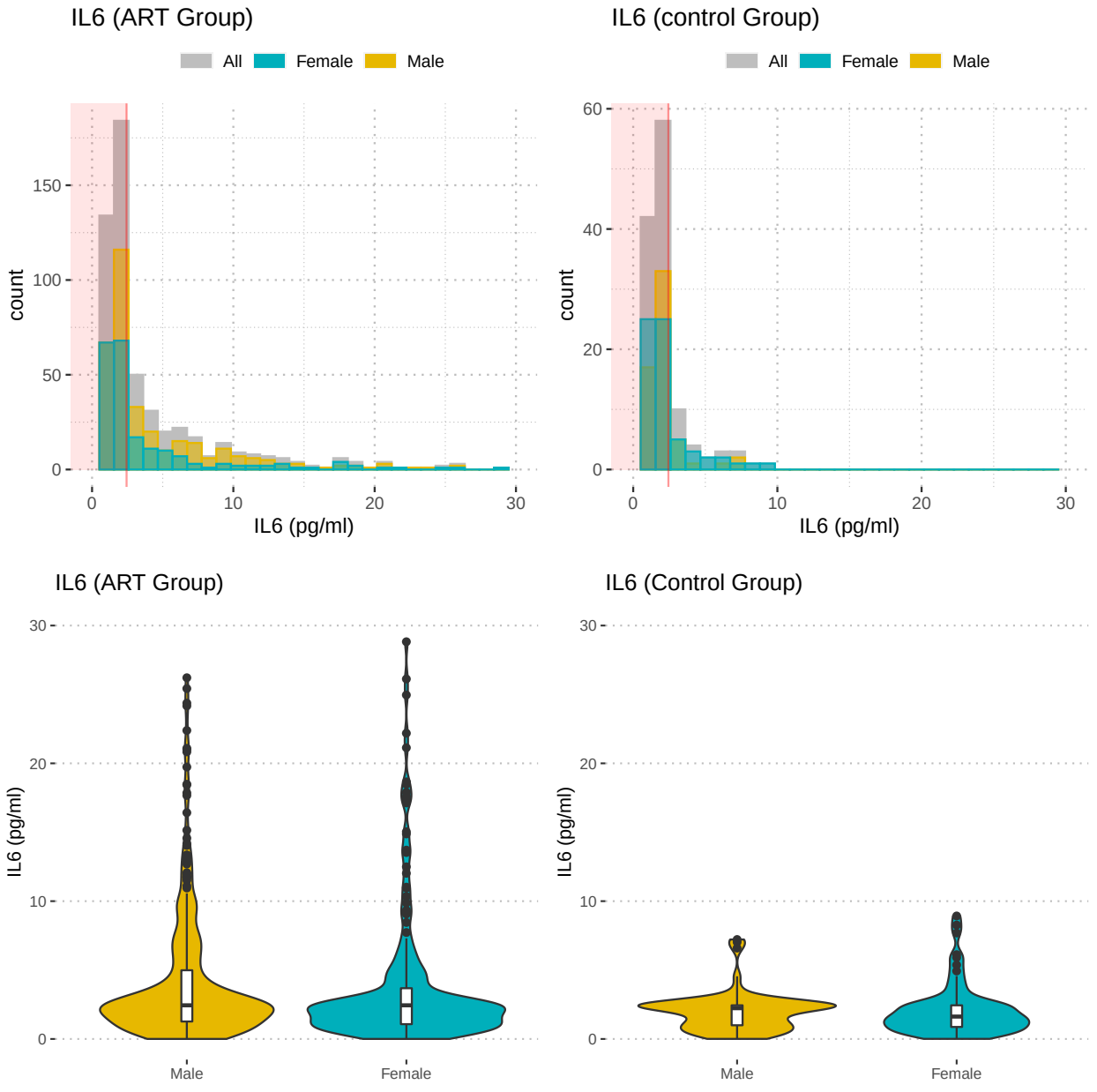


Truncated:



IL6: Sex

We see both genders experiencing significantly higher IL6 levels among ART patients than among control subjects, though the graphs don't seem to show a difference between genders.



IL6: Age

Separating by age group does not seem to show significant differences in distribution