Goodness of fit of measurements from a bioassay (redacted) between customer-site against laboratory distributions

DOCUMENT: SAR-2022-035-SP-v01

From: Felipe Figueiredo **To:** Undisclosed Medical Equipment Company (US)

2022-12-12

TABLE OF CONTENTS

Goodness of fit of measurements from a bioassay (redacted) between customer-site against laboratory distributions

Document version

Version	Alterations
01	Initial version

1 ABBREVIATIONS

- CI: confidence interval
- GoF: goodness of fit
- IQR: interquartile range
- QC: Quality control
- SD: standard deviation

2 CONTEXT

2.1 Objectives

Test the goodness of fit of distribution of measurements from a bioassay (redacted) between laboratory and customer-site conditions.

2.2 Data reception and cleaning

See section 3 (Methods).

3 METHODS

The data procedures, design and analysis methods used in this report are fully described in the annex document **SAP-2022-035-SP-v02**.

4 RESULTS

4.1 Distribution of Analyte measurements

There were 492,648 Analyte measurements included in the dataset, collected from 3 lots from two types of assays: customer-site equipment and reference laboratory assays (Table 1) The majority of measurements available for analysis came from customer equipment, while 186 (<0.1%) were tested in laboratory conditions (Golden = 1).

Table 1 *Origin of Analyte measurements.*

Characteristic	N = 492,648
Lot, n (%)	
0063	147,115 (30%)
0121	181,334 (37%)
0231	164,199 (33%)
Golden, n (%)	
0	492,462 (100%)
1	186 (<0.1%)

The distribution of Analyte measurements across lots and assay conditions is summarized in both Table 2 and Figure 1. Overall, average Analyte are similar across all factors of the design. This is also true for quartiles (including the median Analyte).

The maximum Analyte of each range is similar, except for customer measurements for lot 0121, that exceeds the respective observation from the lab by almost one unit. The minimum values observed at customer locations severely deviate from the reference values, in some cases being almost two units below their respective reference.

These discrepancies give rise to the heavy tails observed in figure 1, especially on the left of each distribution. Note that the plotting strategy of Figure 1 standardizes scales on both axes, for accurate visualization and to allow direct comparison of frequencies. For an alternative version of this figure, cropping each panel to its observed range, see Figure A1 (Appendix).

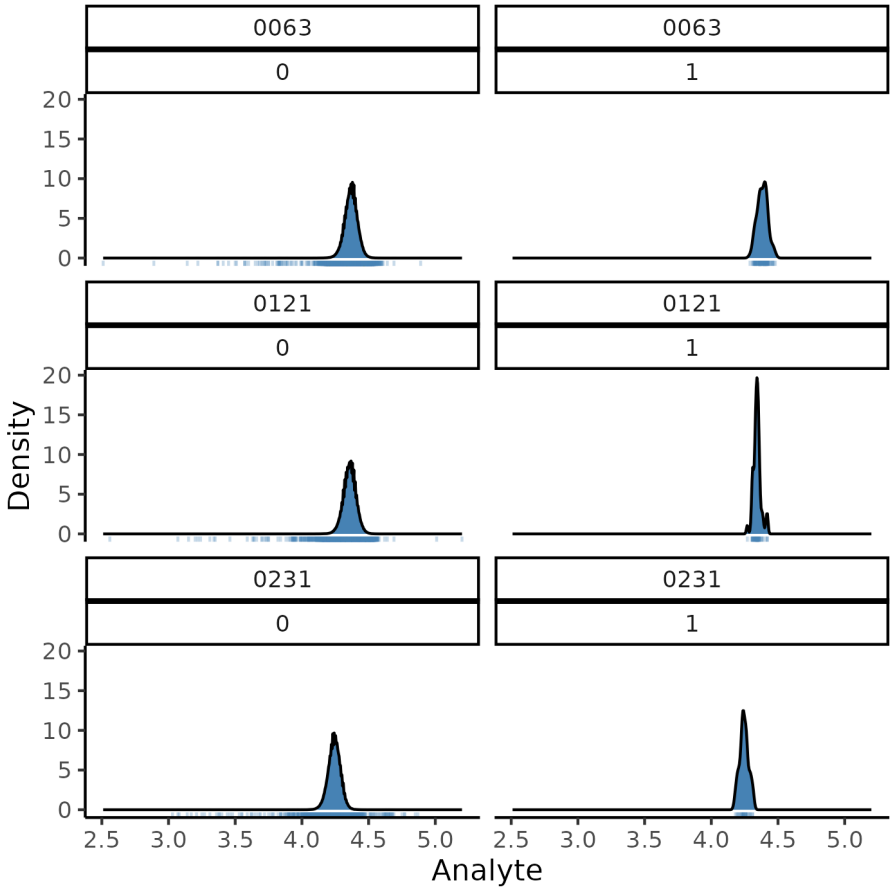

Figure 1 *Density of the distribution of Analyte measurements, by lot and by origin.*

Table 2 *Summary statistics of the distribution of Analyte measurements, by lot and by origin.*

lot_number	golden	Mean	SD	Min	Q1	Median	Q3	Max
0063	0	4.37	0.05	2.51	4.34	4.37	4.40	4.89
0063	1	4.38	0.04	4.29	4.36	4.38	4.41	4.48
0121	0	4.36	0.05	2.56	4.33	4.36	4.39	5.20
0121	1	4.34	0.03	4.27	4.33	4.34	4.35	4.42
0231	0	4.24	0.05	3.03	4.21	4.24	4.27	4.87
0231	1	4.25	0.03	4.18	4.23	4.24	4.26	4.31

4.2 Deviation from normality

Given a strong left tail was identified in all customer site equipment, we'll now investigate its effect on the expectation that these measurements should follow a Normal distribution. Figure 2 is a Q-Q plot of the samples' quantiles against the theoretical quantiles that would be expected if the data was normally distributed.

Overall, in each panel, the lines of both conditions visually match each other, and there is no distinguishable issues at the extremities of the laboratory condition line (strong blue). The deviations from the customer equipment can be inspected more closely here.

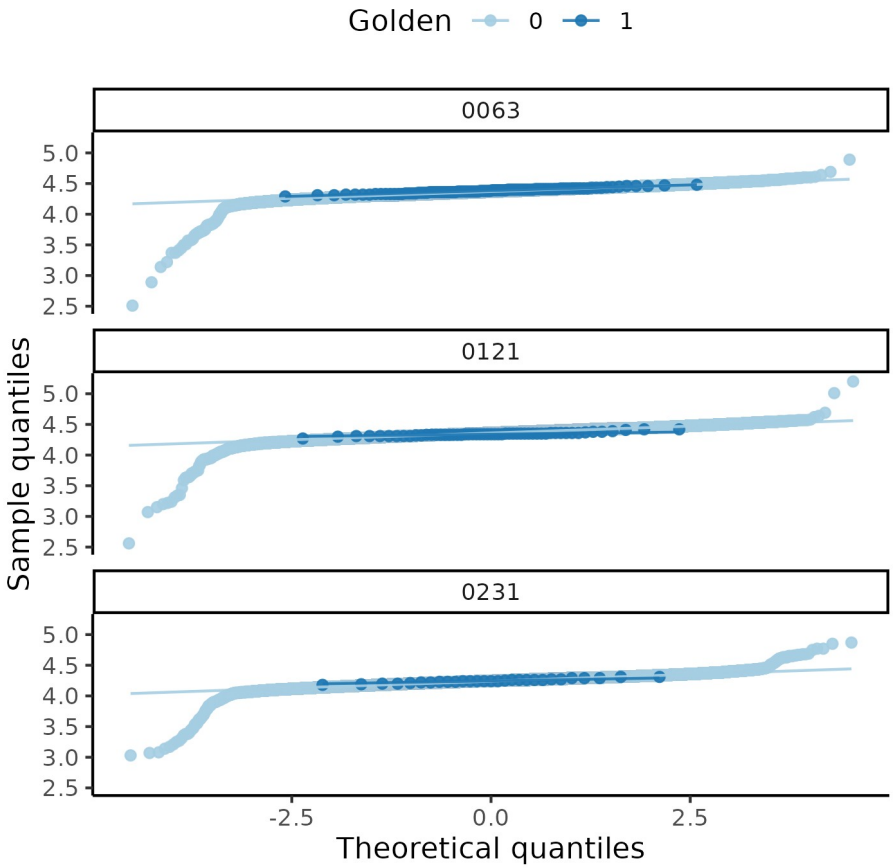

Figure 2 *Q-Q plot of Analyte measurements, by lot and by origin.*

The small deviation on the right extremity is noticeable, but this plot shows it's sufficiently close to the line of perfect match between expectation and observation. The deviation on the left, however, is both more abundant and much more severe, across all conditions.

If these deviations on the left were disregarded, most of the smaller deviations do not appear too far from a normally distributed data set, at least when the laboratory measurements are considered (strong blue). Assuming the data-generating process follows a normal distribution, these deviations could indeed indicate an issue that would warrant intervention from the Quality department. The removal of this data, however, should only be considered under the hypothesis that the measurements were actually failures, either due to equipment malfunction, run-effects of other technical issues. Any such possibility can only be confirmed by equipment technicians or assay experts.

To further investigate this deviation at the left tail of the observed distribution, a complementary perspective can be explored on a scatter plot of the data (Figure 3), by date. This plot shows when the deviations occurred, and by how much they deviated from the average.

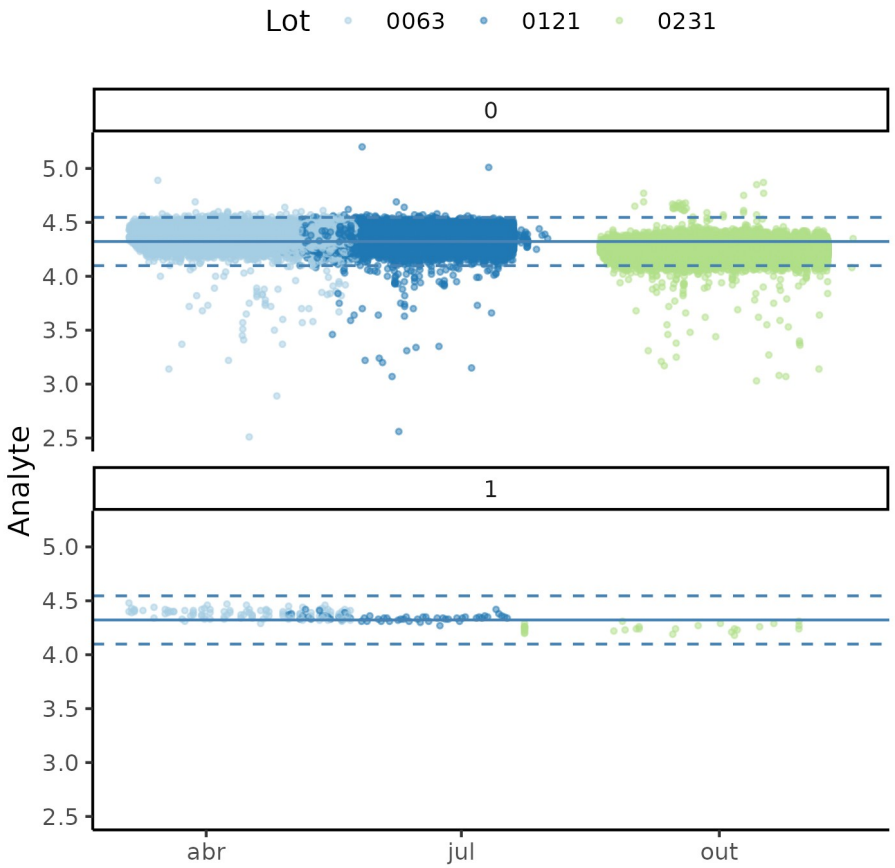

Figure 3 *Scatter plot of Analyte measurements, by lot and by origin. Solid line is the global average of all Analyte values, dashed lines delimit the region between ± 3× SD from the average.*

This plot uses a simple range of mean ± 3×SD, which should enclose 99.7% of expected observations (see Observations). Although this plot is too simplistic for a more detailed QC analysis, it is enough to display the range of the deviations. As seen in the previous section, deviations towards the higher extremity of the analytical range are both rarer and closer to the expected range. The deviations below what would be expected are much more frequent and to a much severe degree.

It is recommended that this visualization strategy be considered in routine QC monitoring for future samples (see Observations).

4.3 Goodness of fit

The GoF between customer-site and laboratory conditions was tested with the asymptotic one-sample Kolmogorov-Smirnov test (Table 2). Customer-site measurements from lots 0063 and 0121 significantly deviate from their respective laboratory measurements, while measurements from lot 0231 do not drift far enough for a deviation to be detected.

It is reasonable to hypothesize that that the data-generating process is following a normal distribution, albeit with some noise being accumulated at the lower end of the analytical range which is being detected here as deviations at the left tail of the distribution.

Table 2 *Kolmogorov-Smirnov test between customer-site and laboratory conditions.*

Lot	p.value
0063	0.009
0121	<0.001
0231	>0.9

In order to visualize these deviations a cropped version of Figure 1 is shown as Figure 4, where the Analyte range was delimited to 3×SD. At this range a serrated distribution of the customer-site measurements becomes apparent. This is compatible with a highly precise assay in which measurements cluster together into tight groups, with gaps between clusters. These clusters of measurements were not apparent in previous visualizations examined so far. A precision profile could be drawn from a validation study to confirm this.

This cropped visualization in Figure 4 may be used to reinforce previous observations that most of the extreme deviations from normality in all lots occur at the left tail of the Analyte distribution. As observed in section 4.2, if those could be discarded the resulting laboratory would be fairly symmetric in lots 0063 and 0231. The serrated distribution of the customer-site measurements already follow a somewhat symmetric distribution,

albeit this observation is solely due to the cropping of the plotting region, and upon inspection of the full analytical this interpretation would be negated by the the extreme values observed on the left.

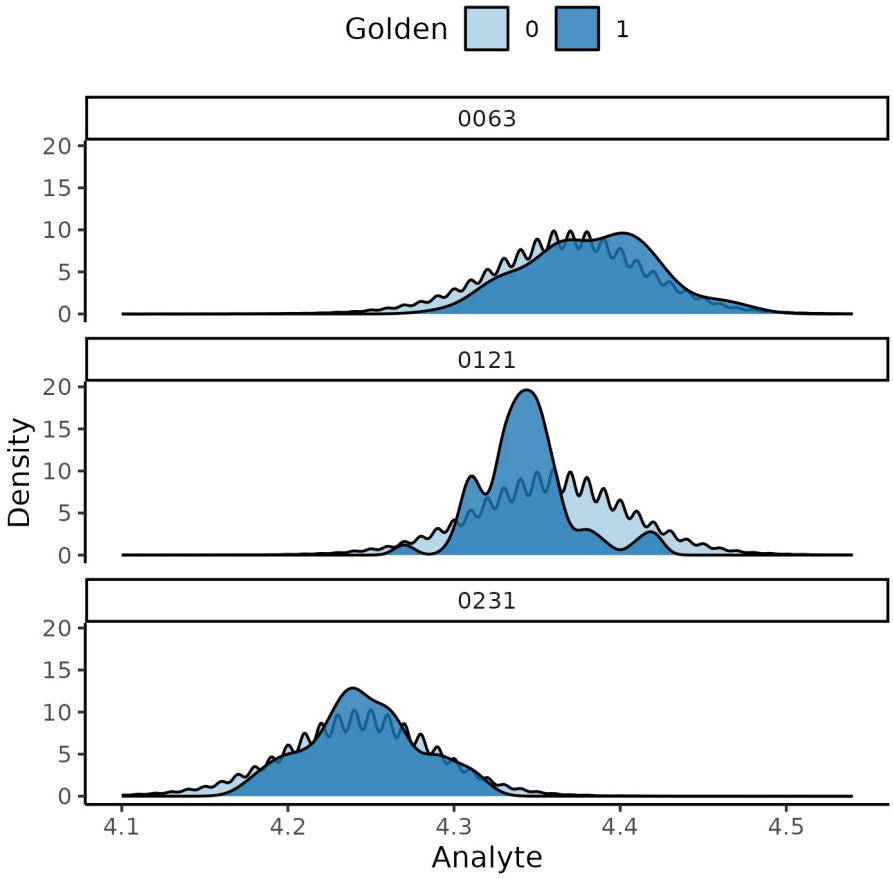

Figure 4 *Distribution of Analyte measurements, by lot and by origin. Plotting range was cropped to ± 3× SD from the average.*

The most unusual laboratory distribution happened at lot 0121, which seems highly multi-modal. This drift could be due to small sample size effects, which could be confirmed in future examination.

5 OBSERVATIONS AND LIMITATIONS

The empirical rule, which states that 99.7% of observations should fall up to 3×SD of the mean, is only valid under the assumption of a Normal distribution.

Furthermore, this statistical principle was used in this report solely as a general guidance for interpretation of the actual data, and not as an acceptance criteria for QC purposes. This rule should not be used for QC decision-making as per the methods in this report.

If the aim of using multiples of the SD is desirable then the ranges between mean and SD should be calculated per lot, and not overall. More precisely, the QC acceptance ranges should be calculated in a context-specific manner, considering different lots, equipment, operators, sites, etc.

6 CONCLUSIONS

- Analyte measurements deviate heavily to the lower end of the distribution, but this happens mostly in customer-site conditions and not under laboratory conditions.
- These deviations make the distribution of customer Analyte fit poorly the reference values, and this happened in all three lots examined.
- If the more extreme deviations could be disregarded as failures, the resulting distribution could realistically closely resemble a Normal distribution, but assessing the causes of those deviations occurred must be confirmed by equipment technicians and assay experts.

7 REFERENCES

- **SAP-2022-035-SP-v02** – Analytical Plan for Goodness of fit of measurements from a bioassay (redacted) between customer-site against laboratory distributions

8 APPENDIX

8.1 Exploratory data analysis

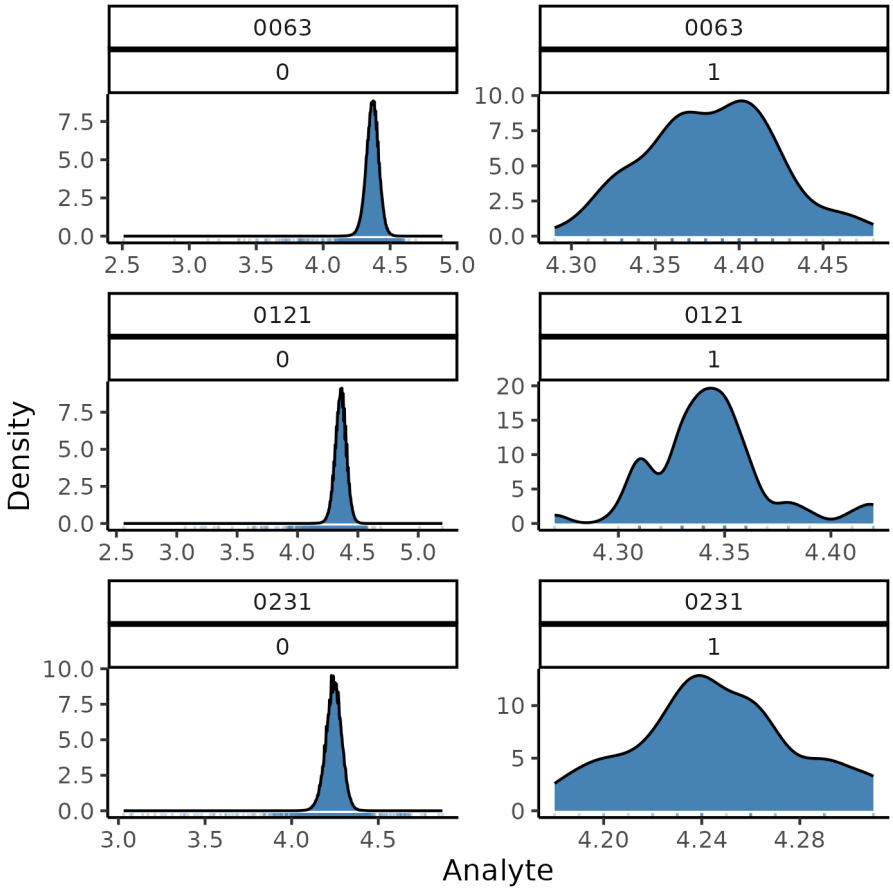

Figure A1 *Alternative version of Figure 1, with free scales on X and Y axes.*

8.2 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

https://philsf-biostat.github.io/SAR-2022-035-SP/

8.3 Analytical dataset

Table A1 shows the structure of the analytical dataset.

Table A1 *Analytical dataset structure*

id	sample_date	lot_number	golden	analyte
1				
2				
3				
…				
492648				

Due to confidentiality the data-set used in this analysis cannot be shared online in the public version of this report.