



Report to participants in the Percentiler- and Flagger programs

Bergen, Dec. 2020

The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus) has produced this report to give you a general overview over some results and findings, mainly from the Percentiler program, but also from the Flagger program.

We want to further develop the programs, and we therefore invite you to send us suggestions for improvements and how we can make the programs more user-friendly.

Please note that Noklus does not permit any reproduction for commercial or other purposes of any part of this report.

We do hope you will find the report useful and look forward to receiving your feedback.

Best regards from

Anne Elisabeth Solsvik (Anne.elisabeth.solsvik@noklus.no)

Program manager

Sverre Sandberg

Director Noklus

Who are currently participating in the Percentiler- and Flagger programs?

At the moment, more than 120 laboratories from 18 different countries are participating in the Percentiler program, and approximately 50 laboratories are also participating in the Flagger program.

The programs were developed in Belgium, and most of the participants are still from this part of Europe. There is wide variation in the laboratories participating, from large private laboratories with many locations, to laboratories at university hospitals and smaller laboratories. It also varies for how many instruments each laboratory reports its daily median values. Some laboratories report an overall median for all outpatients, not one median per instrument.



At the moment you can participate with patient median values and percent flagged results for the following analytes:

ALP	ALT	AST	Bilirubin	BUN
Ca	Cholesterol	Cl	Creatinine	CRP
Ferritin	Folate (B9)	FT4	GGT	Glucose
Hb	HbA1c	HDL-cholesterol	IgA	IgG
IgM	K	LDH	MCV	Mg
Na	Phosphate	PLT	Protein	PSA
PTH	RBC	Triglycerides	TSH	Urea
Uric acid	Vitamin B12	Vitamin D	WBC	

Instruments and grouping

The largest instrument group by far is from Roche and consists of different Cobas models. The second largest group consists of Abbott Architect instruments. To illustrate the number of different instruments participating in the Percentiler program, we counted the number of median values reported from the different instrument groups for three analytes (creatinine, TSH and haemoglobin) a random weekday day in January 2020 (Table 1). This represents an estimate of how many instruments contribute to the monthly median values presented in the diagrams below. Please note that we do not take into consideration the number of patient results from which each median value is calculated, just the number of medians reported.

Table 1 The number of reported median values for three analytes, representing ordinary clinical chemistry analytes (creatinine), immunoassay analytes (TSH) and haematology (haemoglobin) a weekday in January 2020.

Instrument group	Creatinine	TSH	Haemoglobin
Siemens Advia	9	-	-
Siemens Advia (haematology)	-	-	6
Abbott Alinity ci	5	3	-
Abbott Architect	35	29	-
Abbott Cell-Dyn	-	-	2
Siemens Atellica	9	6	-
Beckman AU	7	1	-
Coulter	-	-	1
Siemens Centaur	-	16	-
Roche Cobas	262	99	-
Siemens Dimension	5	3	-
Roche Integra	1	-	-
Roche Modular	3	-	-
Beckman Synchron	4	1	-
Ortho Vitros	17	5	-
Sysmex	-	-	65
Total	357	168	74

Advantages of participating in the Percentiler and Flagger programs

The patient median values reflect both preanalytical and analytical conditions over time. Lack of commutability for both traditional internal quality controls (IQCs) and external quality controls (EQAs) can be a challenge when interpreting results from IQC and EQA. This is overcome when using patient medians. Participation in the Percentiler program can improve your ability to confirm or reject findings detected in IQC and EQA-programs and thereby provides an added value to internal and external quality controls (1, 2). However,

you should bear in mind that composition of the population from which you extract your medians will influence the result.

The main purpose for a laboratory to participate in the programs is to compare its own median with the overall (long term) median from its instrument group, and with the median for all results. From the individual laboratory's point of view, participation in the Percentiler- and Flagger programs should contribute to assess important changes in the long-term median value for an analyte. It can also help to evaluate if a significant change is due to local conditions or caused by the manufacturer.

In addition, the results from a large number of laboratories participating in the programs can provide a basis for assessing the overall harmonization/standardization for different measurement procedures. The website Harmonization.net is a resource centre for information on global activities to harmonize and standardize clinical laboratory measurement procedures. Results from the Percentiler-program can be used to identify problem areas in need for harmonization efforts and assess the effect of established and ongoing harmonization and standardization projects.

The Flagger-program visualizes the effect a possible systematic deviation of your measurement procedure has on the flagging-rate (percentage of results outside the reference limits). Flagger results depend on the concentration of the analyte and reference limits. When comparing your flagging rate to the overall flagging rate you should take into consideration that the reference limits used by laboratories vary.

How to use the programs

User interfaces allow dynamic on-line monitoring of mid-to-long-term stability of performance and flagging rate. It is necessary to get an impression of your own baseline in order to detect important changes. This means that all laboratories should have routines to regularly review results in the Percentiles and Flagger programs.

Tips on how to make the most of your participation in the Percentiler- and Flagger programs

- ❖ Your laboratory procedure manual should include a regular review of your moving medians reported to the Percentiler program. Your results can be a supplement to internal quality control, especially when the control material is not commutable.
- ❖ When you receive deviating EQA-results, especially when the EQA control sample is not commutable, your results in the Percentiler can help you assess if the deviating result is due to local conditions or to a systematic deviation between your measurement procedure (MP) group and other MPs.
- ❖ Participation in the Flagger gives additional information about the consequences of a potential bias for the number of results outside locally used reference limits.

- ❖ Participation in the Percentiler- and Flagger programs can help to assess the effect of reagent or calibrator lot-to-lot variation.
- ❖ Effects of changes in pre-analytical factors, like sample tubes, can be seen in the Percentiler- and Flagger programs.
- ❖ High variation due to the composition of the population in low-throughput laboratories can sometimes be compensated by increasing “n” for calculation of the moving median (the moving median is calculated from minimum 5 daily median values to maximum 16 daily median values).

Performance specifications in the Percentiler- and Flagger programs

Specifications in the Percentiler- and Flagger programs are mainly based on the long-term median value or biological variation, and some target values are based on results from the NORIP study (4).

Noklus plans to revise performance specifications in cooperation with the international expert group that we hope to establish.

The specifications used in the Percentiler- and Flagger are listed in table 2.

The Percentiler

Figure 1 (below) is from the Percentiler-program and presents results and specifications in the program. The upper part of the table with headline **Range: all** refers to data for the complete period and is static also when the range is changed. The lower part of the table with heading **Range: period** refers to data for a selected range and is dynamic.

The table lists:

Value

- Your Your median value
- Peer/All The median value for your peer group or for all results
- Target A fixed value

Bias (%)

- Your bias versus peer group or all results
- The bias for your peer group or all results versus the target
- The proposed acceptance limits

Robust CV (%)

- Your robust CV
- The robust CV for your peer group or for all results

Robust CV is calculated from reported median values.

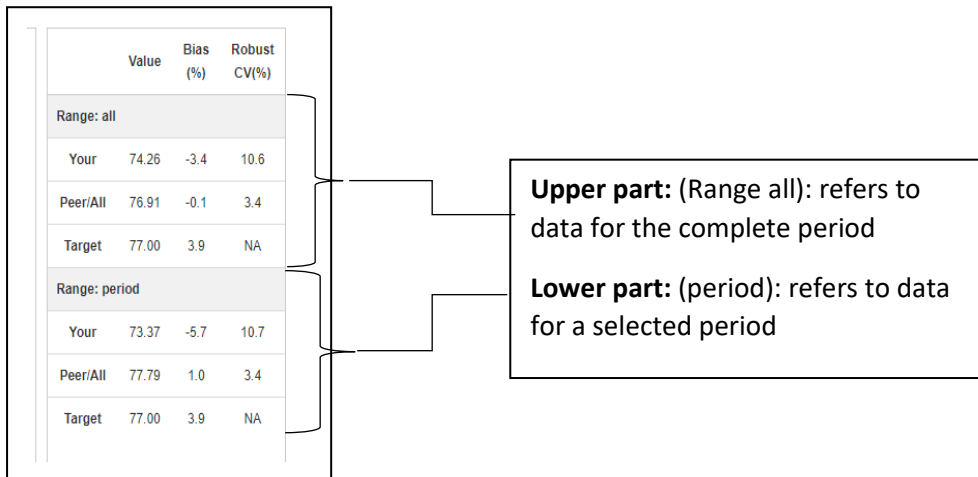


Figure 1 Example from the Percentiler-program for creatinine. The target is set to 77.00 µmol/L. The bias between Peer/All and the target is - 0.1 % based on the results from the complete period reported to the Percentiler (Range: all). For the chosen time period (Range: period), here the last year, the bias between Peer/All and the target is 1.0 %.

Figure 2 (below) shows the diagram presented in the Percentiler-program. The blue full-color line represents the moving median of the selected MP over the selected period. The long-broken grey line shows the long-term median value of the same MP, the short-broken grey lines are the limits for the respective analyte (the area between is shaded). The peer group moving median value or the moving median value for all results is shows as a black broken line.

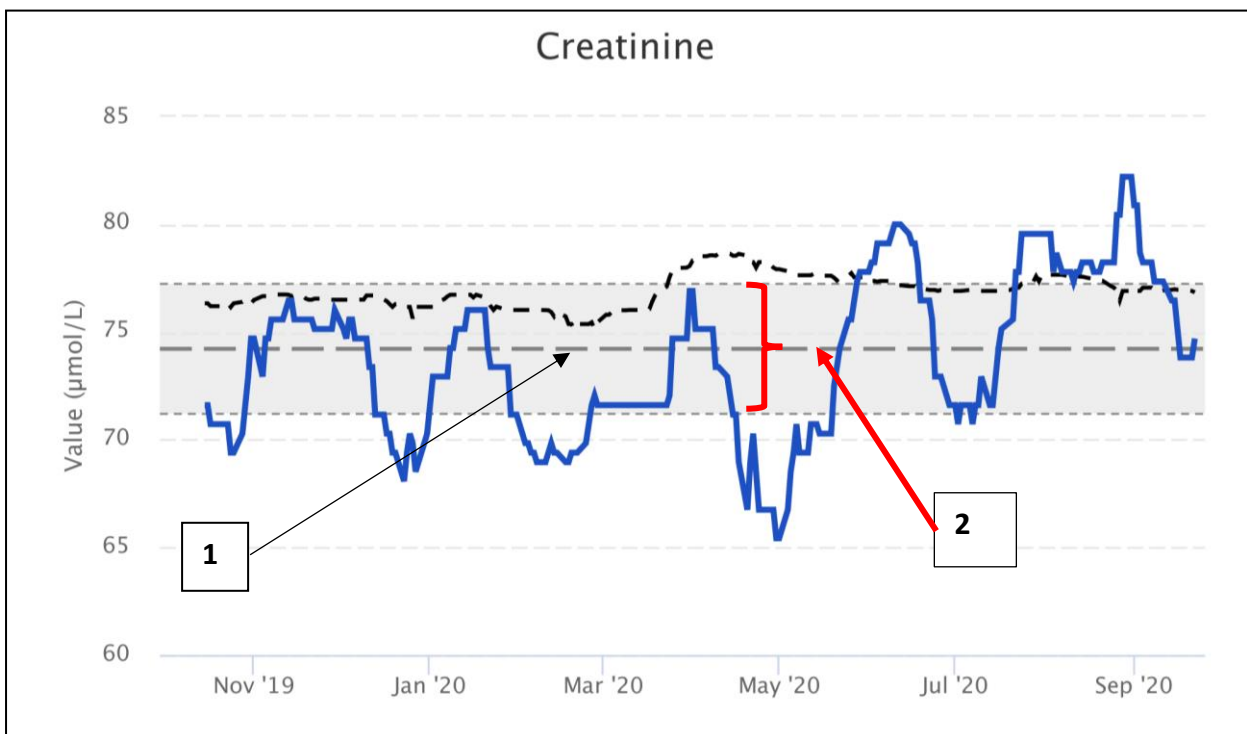


Figure 2. Results for creatinine. The diagram shows: 1) the long-term median for the laboratory (long-broken gray line) at 74.26 µmol/L (Your median value in Figure 1, range all) and 2) the recommended acceptance limits for the respective analyte (the gray shaded area) representing 74.26 µmol/L +/- 3.9%.

Target: The “target value” is found in the literature or the over-all median value in the Percentiler-program. The target value should be the typical median for a healthy population and it is the value to which the **Peer/All** value is compared.

Recommended acceptance limit (see figure 2): The area between the long-term median value for the lab +/- the recommended acceptance limit is the area the moving median value is expected to vary within if the patient population is stable and the number of patient results is sufficient (the grey area around the long-term median value in the Percentiler-diagrams). For creatinine the limit is +/- 3 $\mu\text{mol/L}$. When the target value is 77.00 $\mu\text{mol/L}$ (table 2), this corresponds to an acceptance limits of 3.9 %, as shown in Figure 1 (**Bias (%)** for the **Target** value).

The Flagger

Figure 3 (below) shows the two diagrams presented in the Flagger-program, **Hypo** shows the moving median of the proportion of results (%) below the locally used lower reference limit (also called the hypo flagging rate), and **Hyper** the moving median (%) of the proportion of results above the upper reference limit (or hyper flagging rate). The blue full-color line represents the moving median of the % of flagged results for the selected device over the selected period. The long-broken grey line shows the long-term median value for the laboratory, the short-broken grey lines are the limits for the respective analyte (the area between is shaded). The peer group moving median value or the moving median value for all results is shown as a black broken line.

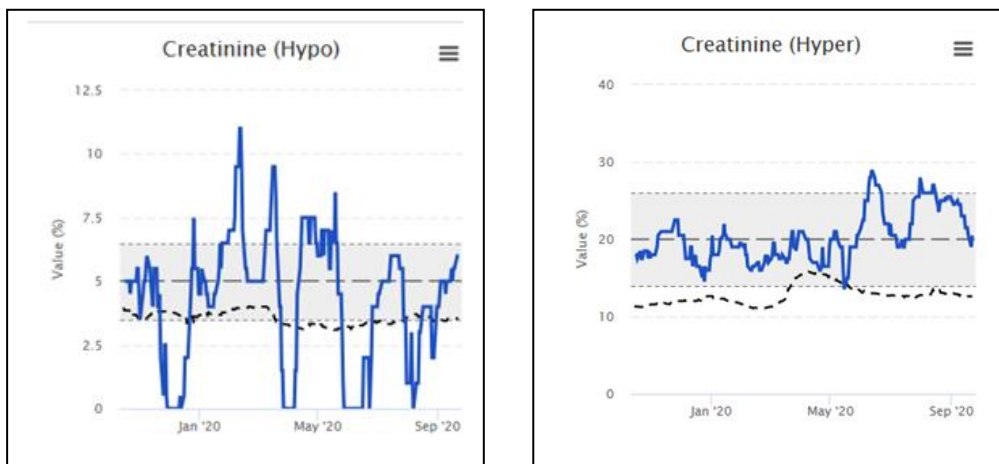


Figure 3. Here are creatinine results as shown in the Flagger for the same lab as in Figure 1 and Figure 2. The long-term flagging rate for this instrument is 5% for results below the low reference limits (Hypo), and 20% for results over the upper reference limits (Hyper). The acceptance limit is 30% Table 1), and the acceptance area in the Creatinine (Hypo) diagram is 5% +/- 1.5% (0,3*5) and for the Creatinine (Hyper) diagram the acceptance area is 20% +/- 6% (0,3*20).

Limit (%): Acceptance limits are indicated with a horizontal, grey dotted line in the Flagger-diagrams (Figure 3). Flagger limits are chosen so that they agree with the current “state-of-the-art” performance for each individual laboratory and these limits are calculated relatively to the long-term flagging rate. For example: if the long-term flagging rate is 10% and the limit (%) is 30 (table 2), the acceptance limits are +/- 3%.

Minimum Limit: The acceptance limit is never set below +/- 1%.

Interpretation of short-term variation should always be handled with caution when the population variation is high, or the population is small.

Table 1. Specifications for analytes in the Percentiler- and Flagger today

Analyte	Unit	Specifications in the Percentiler		Specifications in the Flagger
		Recommended acceptance limit	Target	Limit (%)
25-hydroxy vitamin D	nmol/L	3	60	50
Albumin	g/L	1	42	50
Alkaline Phosphatase	U/L	5	73	30
ALT	U/L	2	21	30
AST	U/L	1,5	23	30
BUN	mg/dL	1	14	30
Calcium	mmol/L	0,04	2,35	70
Chloride	mmol/L	1	102,5	50
Creatinine	µmol/L	3	77	30
CRP	mg/L	0,25	2,6	30
Ferritin	pmol/L	8	157	50
Folate	nmol/L	1,5	14,4	50
FT4	pmol/L	0,5	15	30
GGT	U/L	2	22	30
Glucose	mmol/L	0,15	4,9	20
Glycated hemoglobin	mmol/mol	1	42	30
Glycated hemoglobin (%)	%	0,3	6	50
HDL-cholesterol	mmol/L	0,08	1,3	30
Hemoglobin	mmol/L	0,2	8,87	50
Immunglobulin A	g/L	0,25	2,4	50
Immunglobulin G	g/L	0,5	11	50
Immunglobulin M	g/L	0,15	1,35	50

Inorganic phosphor	mmol/L	0,05	1,13	50
LDH	U/L	8	175	30
Magnesium	mmol/L	0,02	0,83	70
MCV	fL	1	90	50
MPV	fL	0,25	9,5	50
Platelets	10 ⁹ /L	15	250	50
Potassium	mmol/L	0,1	4,25	30
PSA	ng/mL	0,75	1,7	70
PTH	pmol/L	1	5	70
RBC	10 ¹² /L	0,1	4,7	50
Sodium	mmol/L	1	140,5	70
Total-bilirubin	μmol/L	1	10	30
Total-cholesterol	mmol/L	0,2	5,2	20
Total-protein	g/L	1	72	50
Triglycerides	mmol/L	0,1	1,4	30
TSH	mIU/L	0,12	1,56	30
Urea	mmol/L	0,3	5	30
Uric Acid	μmol/L	15	310	50
Vitamin B12	pmol/L	30	300	50
WBC	10 ⁹ /L	0,4	5,7	50

Future plans

Noklus plans to further develop both programs. In short, depending on available resources, Noklus will:

1. Establish an international advisory board.
2. Develop a new database and user interface for the Percentiler and Flagger and establish and implement new performance specifications.
3. In cooperation with EQA providers make proposals for how traditional EQA can be combined with the Percentiler and Flagger-programs.

We hope that many of you will contribute to developing even better programs for patient medians!

Results

We present results from the Percentiler and Flagger -programs for several analytes for the time period August 1st 2019 through July 31th 2020. First you find results for selected analytes with a comment added, followed by analytes without comments.

Results from the Percentiler-program

In the figures below, for each analyte you can see the monthly median per instrument group during August 1st 2019 – July 31th 2020. The smallest instrument groups are excluded.

In the tables, the median value is the median of the monthly medians for the specific instrument group during August 1st 2019 - July 31th 2020. Weekend results are excluded. The minimum value (min) is the median value for the month with the lowest monthly median. The maximum (max) value is the median value for the month with the highest monthly median. The tables also include the average number of median values reported per month for each instrument group.

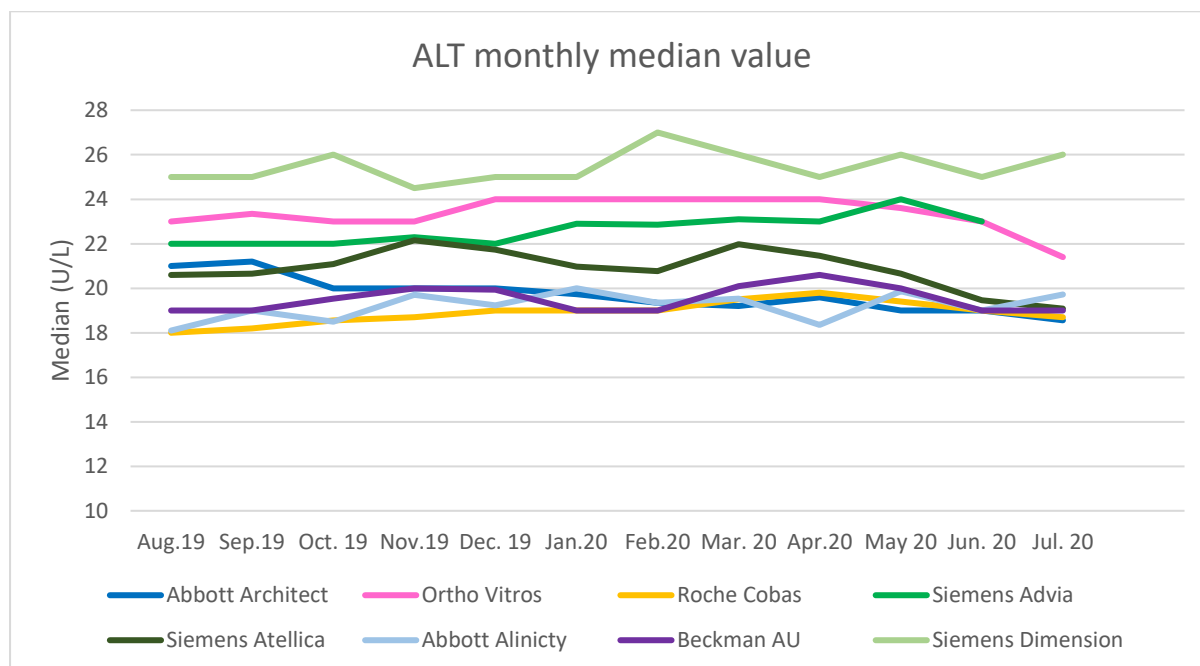
Results from the Flagger-program

For selected analytes there is also a table presenting results from the Flagger-program, where the median value calculated from results for the specific instrument groups during August 1st 2019 - July 31th 2020 is presented. In addition, the 25- and 75-percentile value is included to give an impression of the variation in the data. You will also find the number of results from which the median value is calculated.

There are also figures presenting the monthly median value per instrument group.

ALT

Percentiler results:



Instrument group	ALT - Median of medians (U/L) (min-max)	N reported medians (pr. month)
Abbott Alinity	19 (18-20)	136
Abbott Architect	20 (19-21)	774
Beckman AU	19 (21-24)	163
Ortho Vitros	23 (23-24)	297
Roche Cobas	19 (18-20)	4723
Siemens Atellica	21 (19-22)	174
Siemens Advia	23 (22-24)	93
Siemens Dimension	25 (25-27)	108

Flagger results – results above the upper reference limit:

Instrument group	ALT - Median value (%) (25- -75 percentile) Results above the upper reference limit	N reported medians (average pr. month)
Abbott Alinity	11 (8-15)	102
Abbott Architect	15 (10-18)	472
Beckman AU	11 (8-14)	140
Ortho Vitros	11 (0-15)	42
Roche Cobas	9 (8-11)	3544
Siemens Atellica	15 (13-18)	103
Siemens Advia	16 (14-17)	64
Siemens Dimension	6 (4-9)	72

Comments:

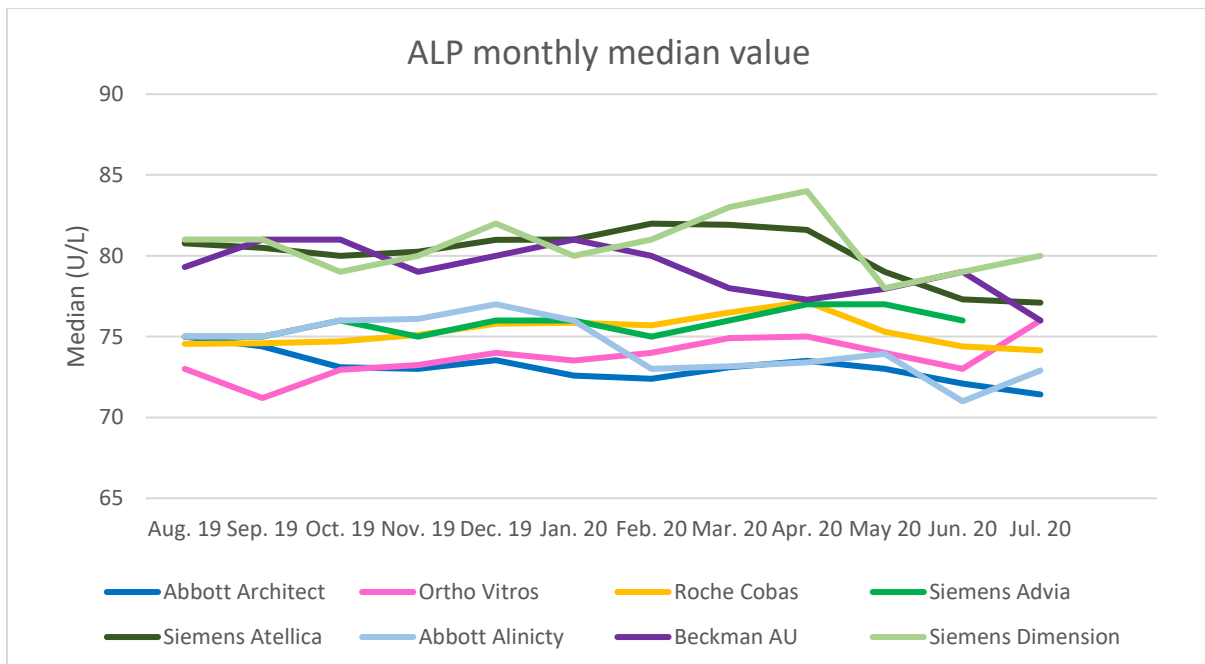
Abbott instruments (Architect and Alinity), Roche instruments (Cobas) and Beckman AU instruments report similar median values. Siemens Dimension and Ortho Vitros have the highest median values, while Siemens Atellica and Advia laboratories also reports higher patient medians than the Roche and Abbott instruments.

EQA results from Labquality also show that both Siemens Dimension and Siemens Advia have slightly higher results than Abbott and Roche instruments.

A Canadian survey from 2017 found similar differences between instruments from different vendors for ALT when a commutable reference sample were analysed in different laboratories and instruments (5).

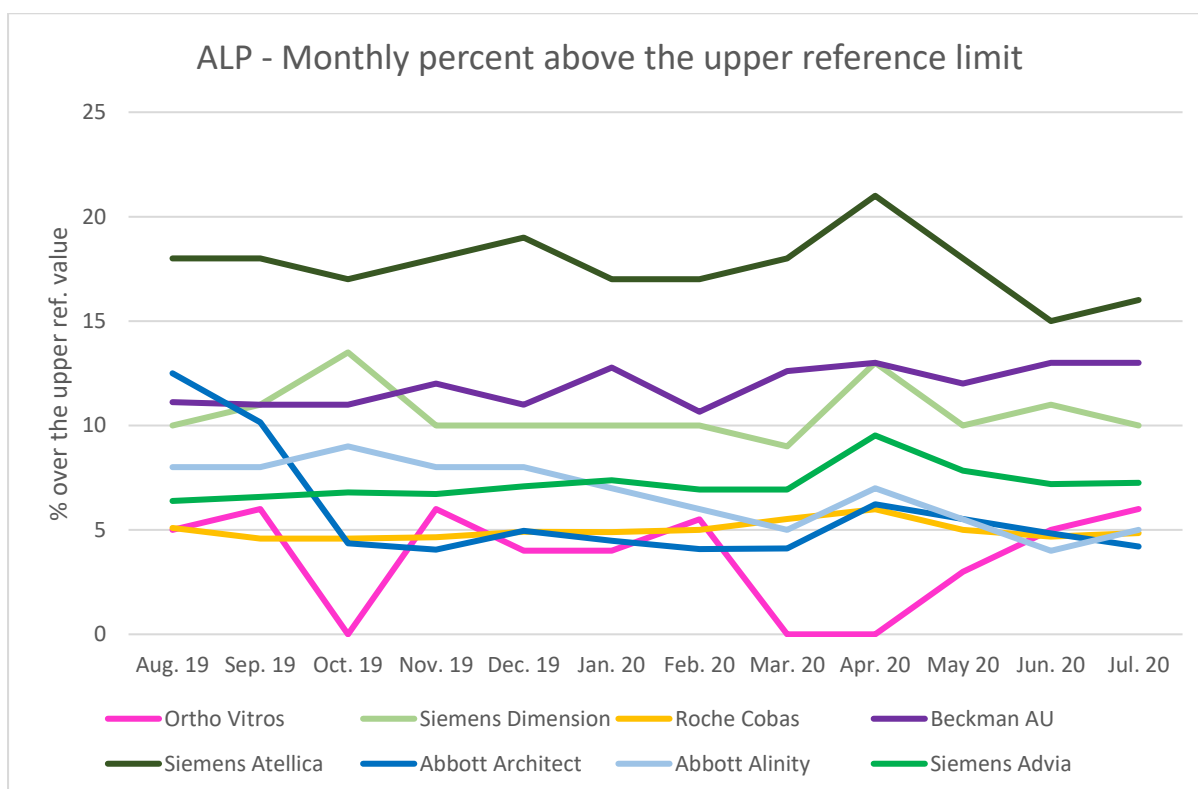
ALP

Percentiler results:



Instrument group	ALP - Median of medians (U/L) (min-max)	N reported medians (pr. month)
Abbott Alinity	74 (71-77)	112
Abbott Architect	73 (71-75)	739
Beckman AU	79 (76-81)	160
Ortho Vitros	75 (71-76)	346
Roche Cobas	75 (74-77)	4620
Siemens Atellica	81 (77-82)	211
Siemens Advia	76 (75-77)	76
Siemens Dimension	81 (78-84)	107

Flagger results – percent above the upper reference limit:



Instrument group	ALP - Median value (%) (25- -75 percentile) Results above the upper reference limit	N reported medians (average pr. month)
Abbott Alinity	7.0 (4.0-12.0)	98
Abbott Architect	5.1 (2.9-11.0)	517
Beckman AU	12.0 (8.0-16.2)	140
Ortho Vitros	4.0 (0.0-10.0)	65
Roche Cobas	5.0 (4.0-6.5)	3539
Siemens Atellica	17.0 (13.0-22.0)	155
Siemens Advia	7.1 (6.3-8.1)	64
Siemens Dimension	10.0 (7.0-14.0)	65

Comments:

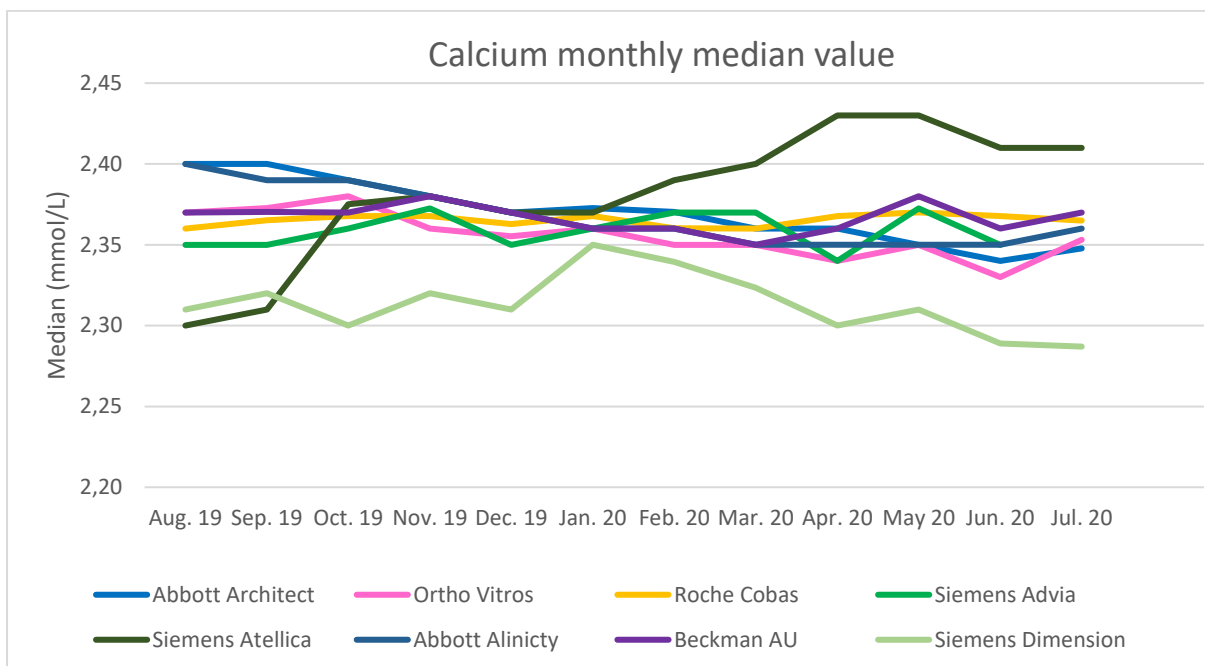
A new standardized measurement procedure for ALP was published in 2011 (6).

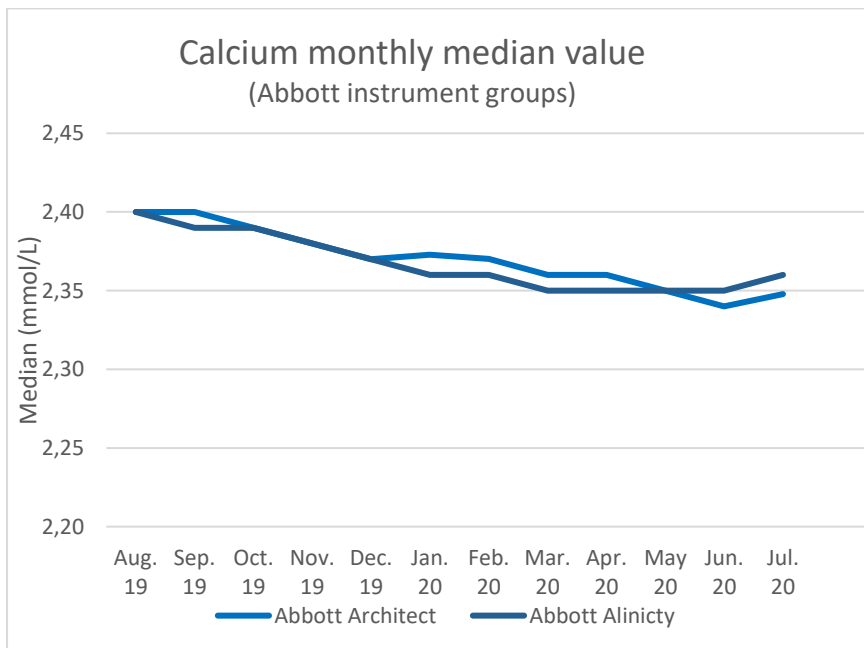
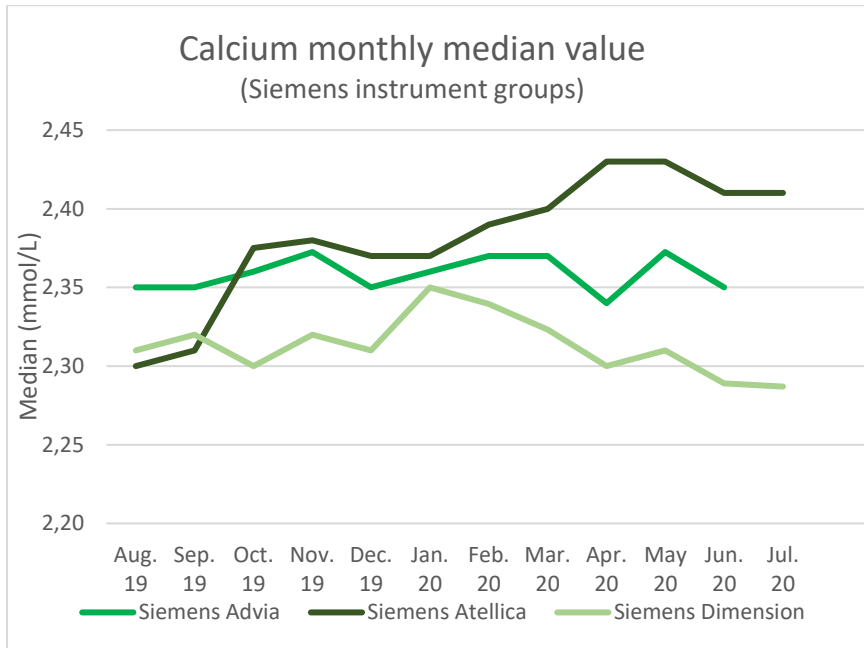
Beckman AU Instruments report a high median value, this is also found in a trueness project evaluating the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories (7).

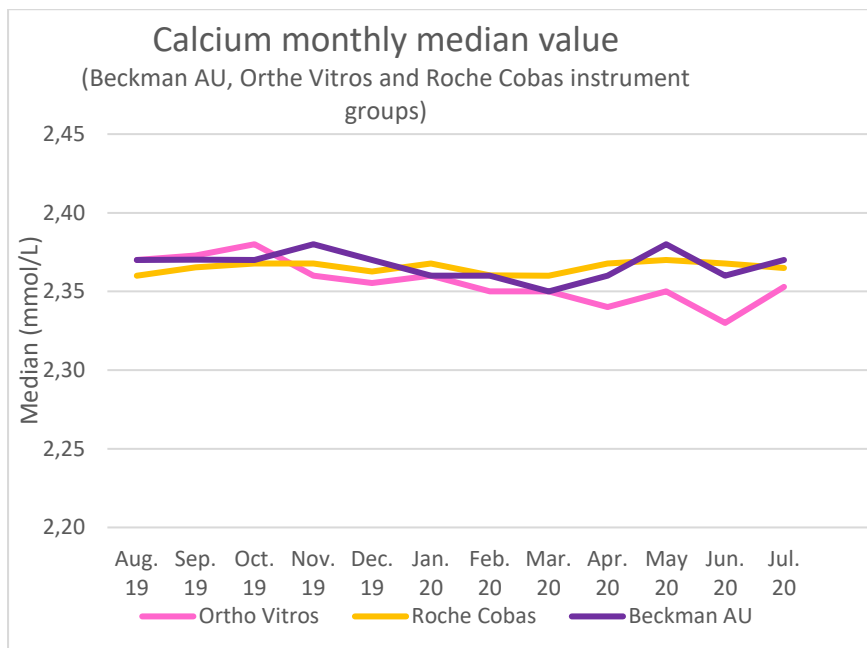
The instrument vendors often offer different methods to their customers, this makes the interpretation more complicated. Results from the Percentiler-program can be used to follow up the implementation of the standardized measurement procedure published in 2011 and if the instrument vendors eventually achieve harmonization of the analyte.

Calcium (Ca)

Percentiler results:

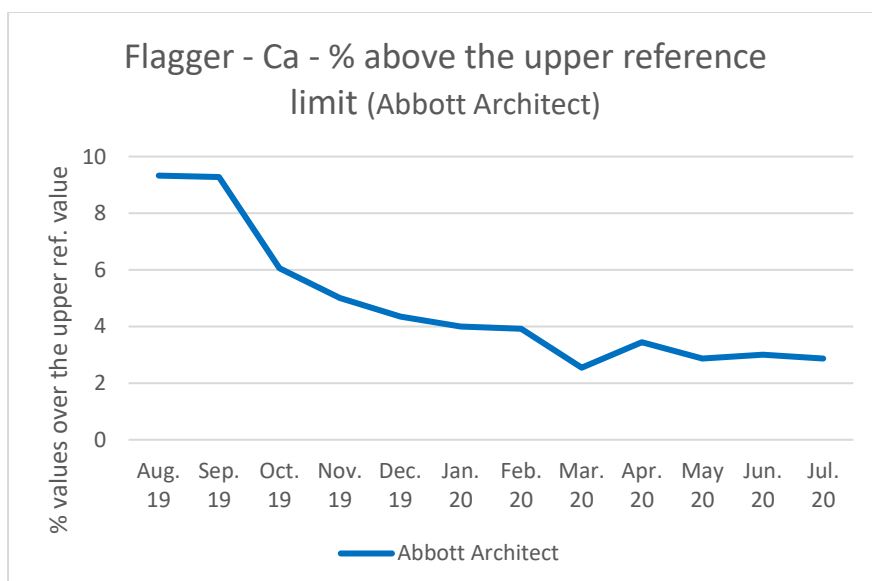


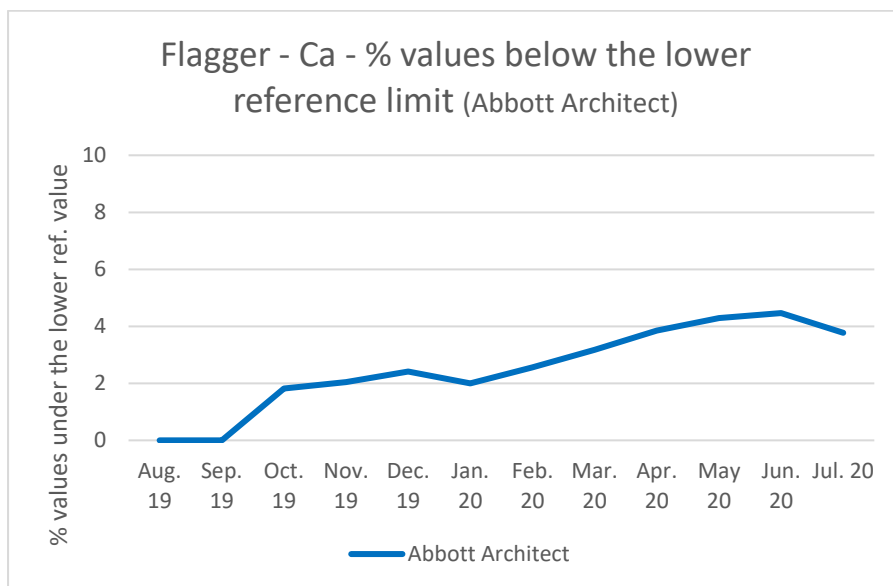




Instrument group	Ca - Median of medians (min-max)	N reported medians (pr. month)
Abbott Alinity	2.36 (2.36 – 2.37)	262
Abbott Architect	2.37 (2.34 – 2.40)	684
Beckman AU	2.37 (2.35 – 2.38)	158
Ortho Vitros	2.35 (2.33 – 2.38)	345
Roche Cobas	2.37 (2.36 – 2.37)	4637
Siemens Atellica	2.39 (2.30 – 2.43)	244
Siemens Advia	2.36 (2.34 – 2.37)	125
Siemens Dimension	2.31 (2.29 – 2.35)	106

Flagger results:





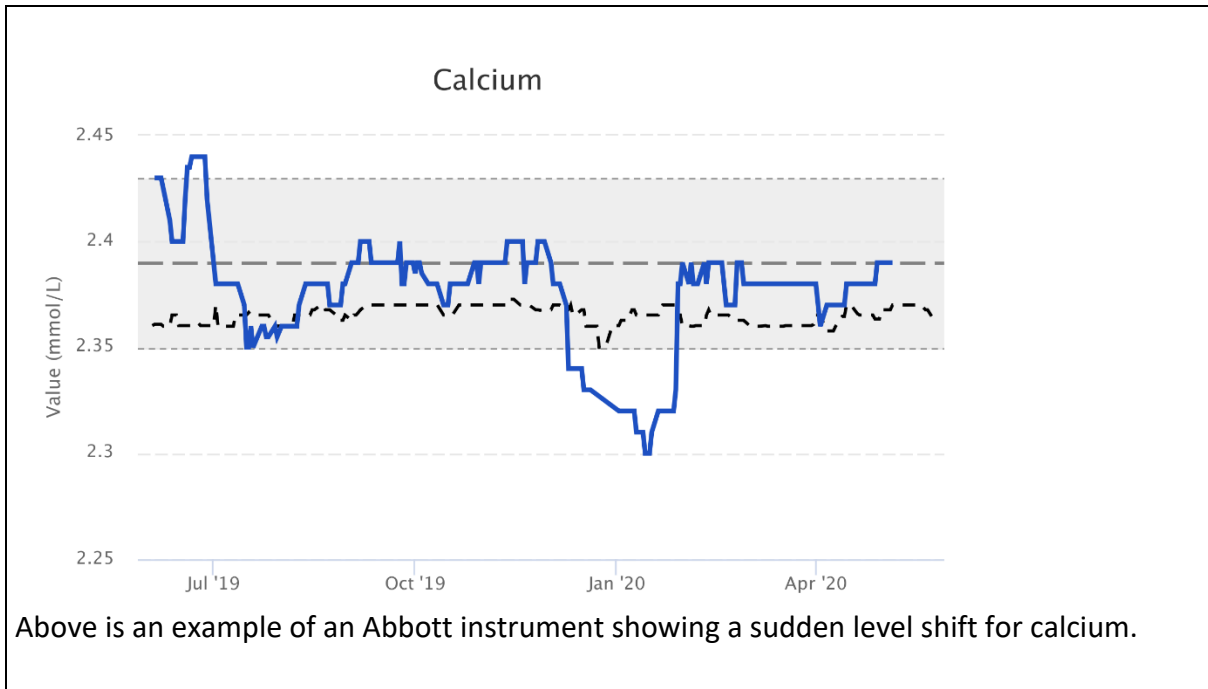
Ca	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)	
	Instrument group	Results above the upper reference limit		Results below the lower reference limit
	Abbott Alinity	3.0 (0.0-6.0)	2.0 (0.0-4.0)	74
	Abbott Architect	4.3 (1.4-9.1)	2.3 (0.0-5.0)	480
	Beckman AU	0.0 (0.0-2.4)	4.0 (0.0-9.0)	138
	Ortho Vitros	7.0 (0.0-14.0)	0.0 (0.0-6.0)	64
	Roche Cobas	5.0 (3.8-6.8)	3.4 (2.4-4.8)	3597
	Siemens Atellica	2.0 (0.0-6.0)	3.0 (0.0-5.0)	156
	Siemens Advia	1.6 (0.8-2.5)	1.1 (0.4-1.9)	64
	Siemens Dimension	0.0 (0.0-3.0)	2.0 (0.0-7.0)	64

Comments:

Perhaps the most striking finding for calcium is the drop for both the Architect and the Alinity peer group during Autumn 2019. This has a big impact on the results in the Flagger (only shown for Architect due to few results for Alinity). We have also seen that several Abbott users have had sudden drops in concentration illustrated in the example below.

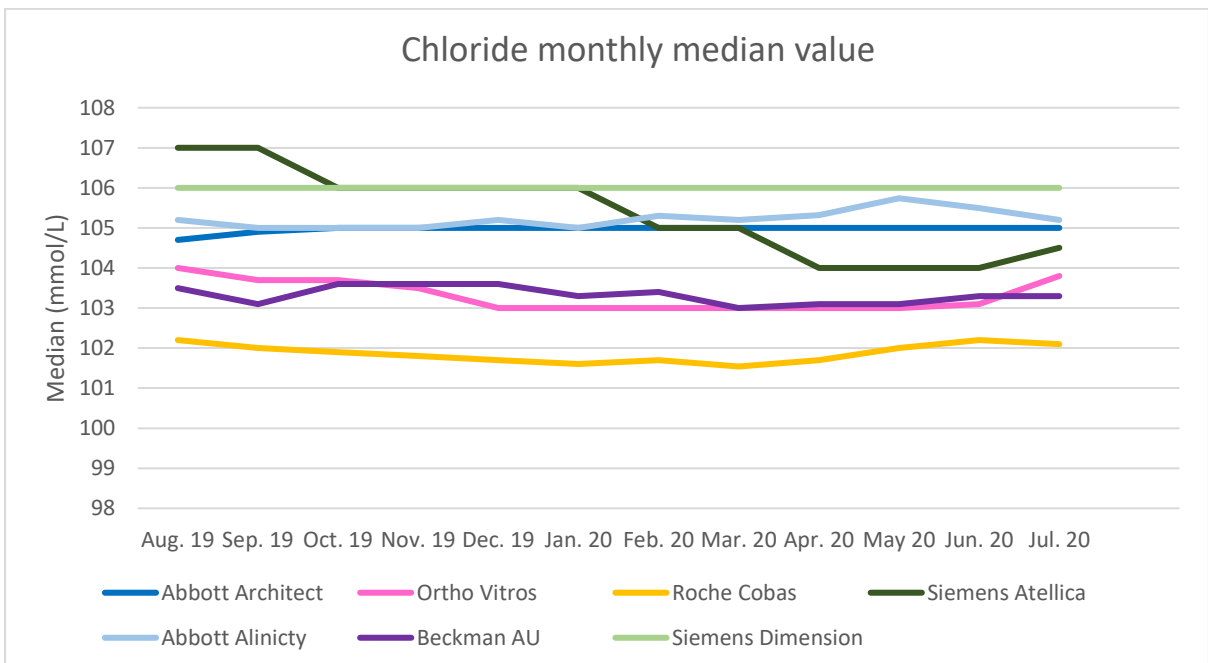
The Siemens Atellica group has the highest median of the monthly medians value, but with wide variation between the lowest and highest monthly median for the group. The Siemens Dimension group reports the lowest monthly median value, also here there is a wide variation between the lowest monthly average median and the highest value.

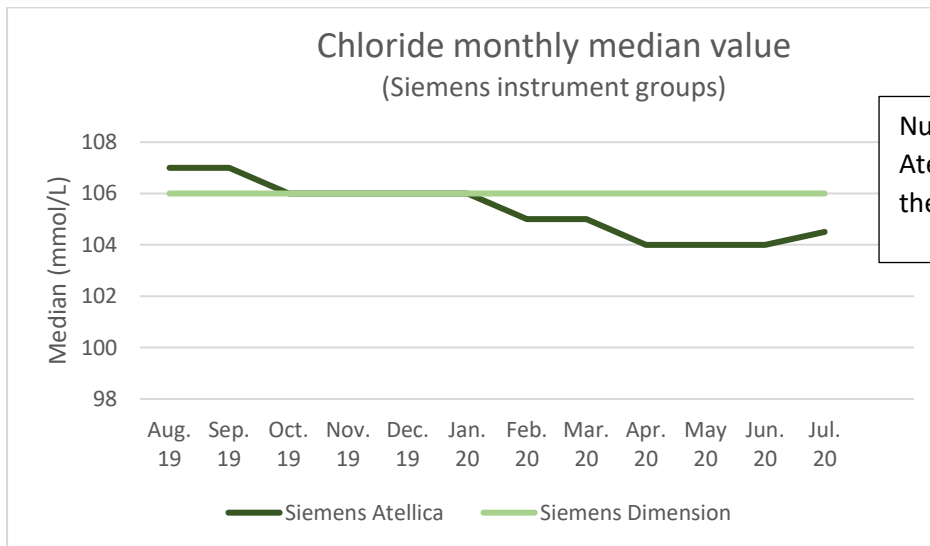
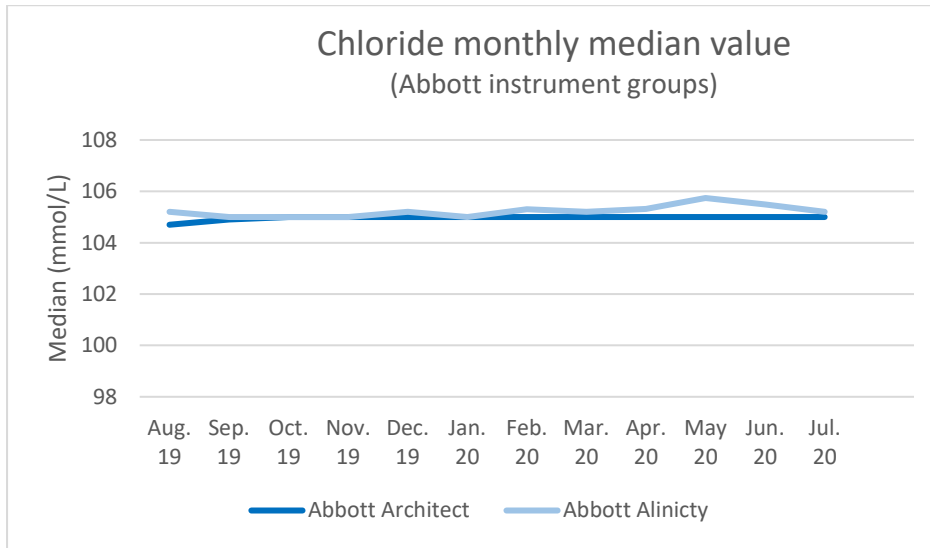
Several of the instrument vendors offer different methods.



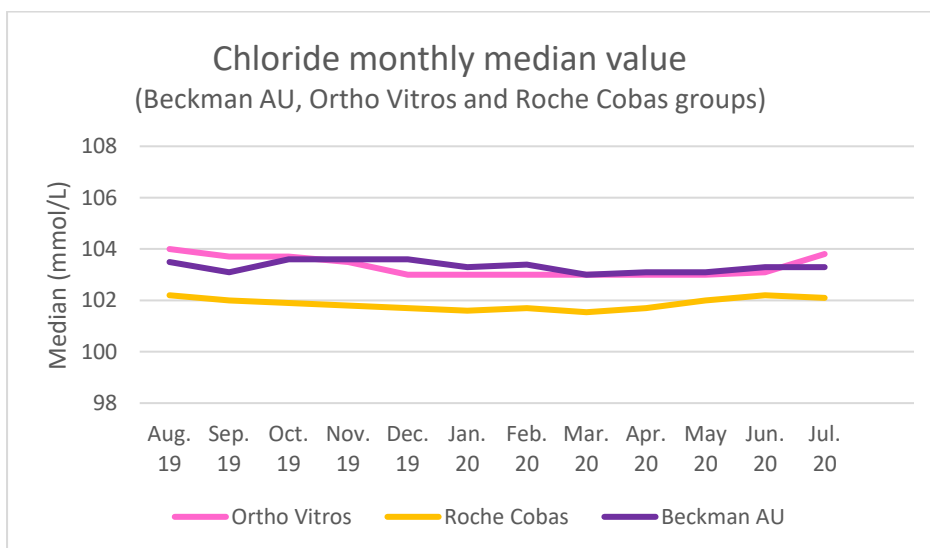
Chloride

Percentiler results:





Number of instruments in the Atellica has doubled during the year.



Instrument group	Cl - Median of medians (mmol/L) (min-max)	N reported medians (pr. month)
Abbott Alinity	105.2 (105.0 – 105.7)	149
Abbott Architect	105.0 (104.7 – 105.0)	615
Beckman AU	103.3 (103.0 – 103.6)	148
Ortho Vitros	103.1 (103.0 – 104.0)	348
Roche Cobas	101.9 (101.5 – 102.2)	4446
Siemens Atellica	105.5 (104.0 – 107.0)	135
Siemens Dimension	106.0 (106.0 – 106.0)	108

Flagger results:

Cl	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
Instrument group	Results above the upper reference limit	Results below the lower reference limit	
Abbott Alinity	17.0 (12.0-23.0)	2.0 (0.0-4.0)	170
Abbott Architect	2.1 (0.0-4.8)	3.2 (1.1-5.8)	392
Beckman AU	0.0 (0.0-2.0)	15.0 (1.1-24.0)	131
Ortho Vitros	7.0 (0.0 – 15.0)	3.0 (0.0-8.0)	65
Roche Cobas	1.2 (0.4-2.6)	7.0 (4.8-10.3)	3146
Siemens Atellica	14.0 (0.0-50.0)	0.0 (0.0-20.0)	65
Siemens Dimension	26.0 (9.0-38.0)	2.0 (0.0-5.0)	65

Comments:

The Cobas instrument group reports the lowest median values for Chloride. Both Abbott instrument groups reports similar medians, and users of the Siemens instrument report the highest patient medians.

This difference between instruments is also seen in EQA from Labquality.

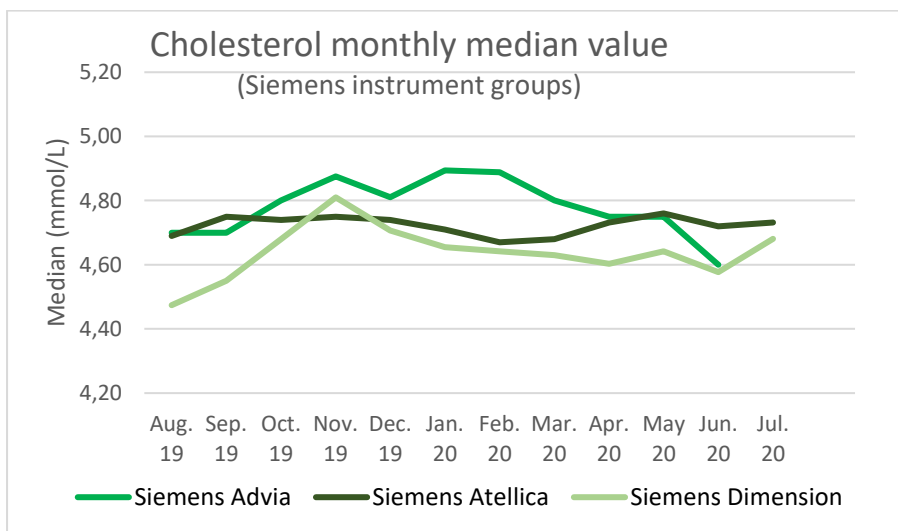
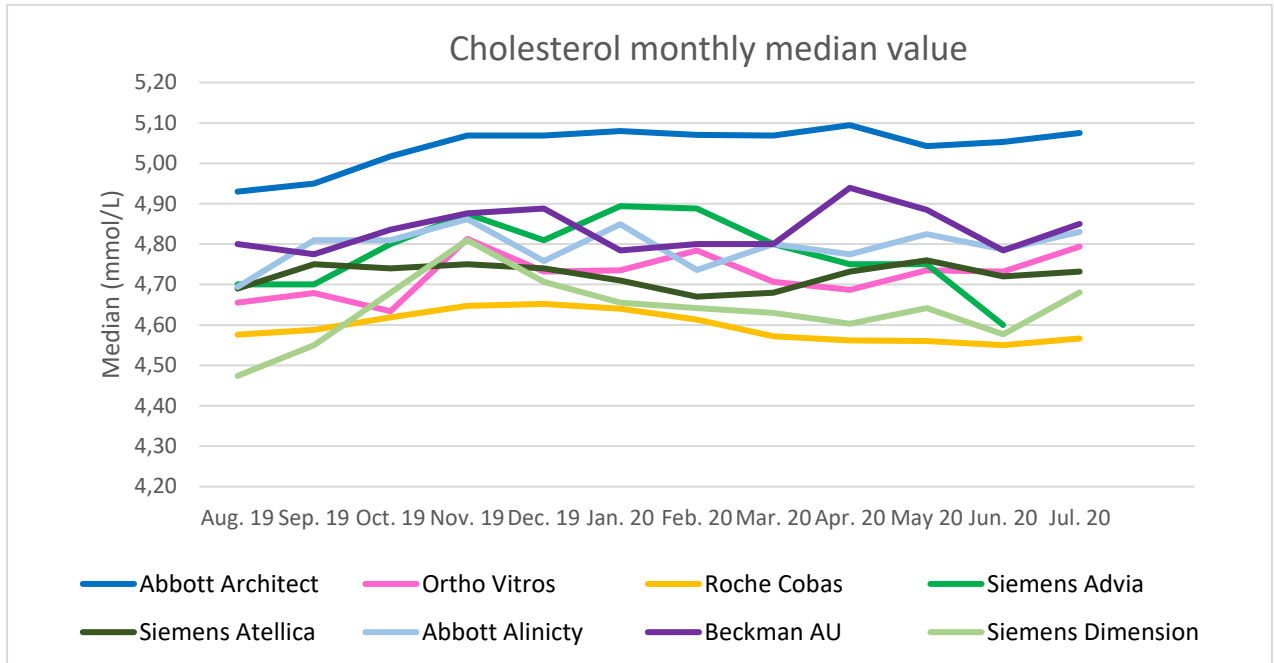
We believe that these results reflect real method differences and point to a metrological difference (best to describe as difference in calibration basis) between the instruments.

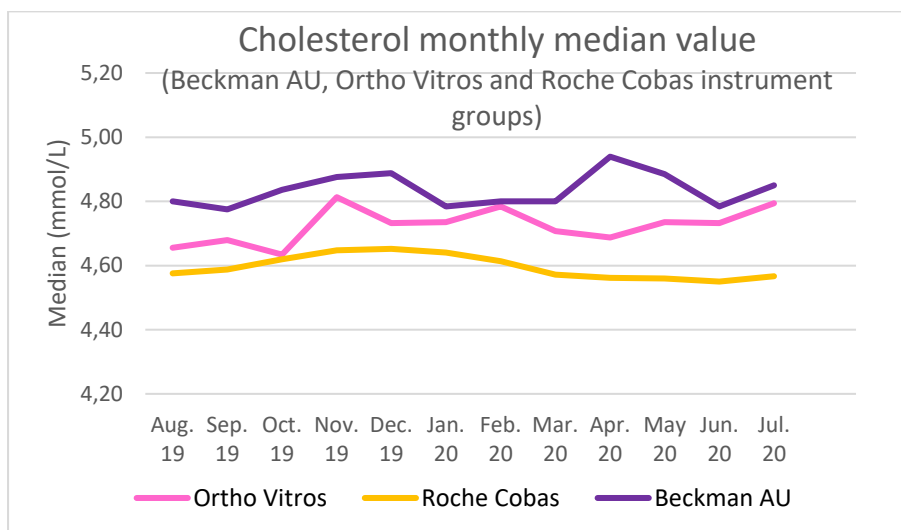
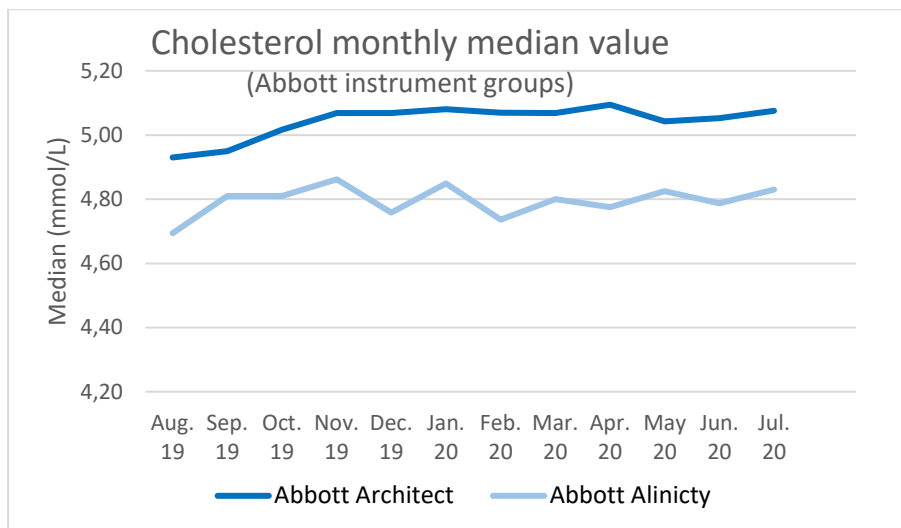
This difference could easily be resolved by harmonization between companies.

Note that the observed differences are clinically relevant.

Cholesterol (total)

Percentiler results:





Instrument group	Cholesterol - Median of medians (min-max)	N reported medians (pr. month)
Abbott Alinity	4.81 (4.69 – 4.86)	111
Abbott Architect	5.07 (4.93 – 5.09)	760
Beckman AU	4.82 (4.78 – 4.94)	157
Ortho Vitros	4.73 (4.63 – 4.81)	294
Roche Cobas	4.58 (4.55 – 4.65)	4548
Siemens Atellica	4.66 (4.47 – 4.55)	158
Siemens Advia	4.60 (4.60 – 4.60)	86
Siemens Dimension	4.58 (4.47 – 4.55)	103

Flagger results:

Cholesterol	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
	Instrument group	Results above the upper reference limit	
Abbott Alinity	20.0 (0.0-46.0)	0.0 (0.0-13.3)	77
Abbott Architect	53.3 (38.3-59.2)	0.0 (0.0-0.5)	524
Beckman AU	41.7 (29.0-40.6)	0.0 (0.0-2.0)	138
Ortho Vitros	35.0 (20.0-50.0)	0.0 (0.0-0.0)	62
Roche Cobas	30.3 (27.1-33.7)	1.4 (0.7-2.0)	3467
Siemens Atellica	8.0 (5.0-23.0)	0.0 (0.0-0.0)	120
Siemens Advia	11.5 (9.9-13.3)	0.0 (0.0-0.0)	64
Siemens Dimension	31.0 (24.0-39.0)	0.0 (0.0-0.0)	64

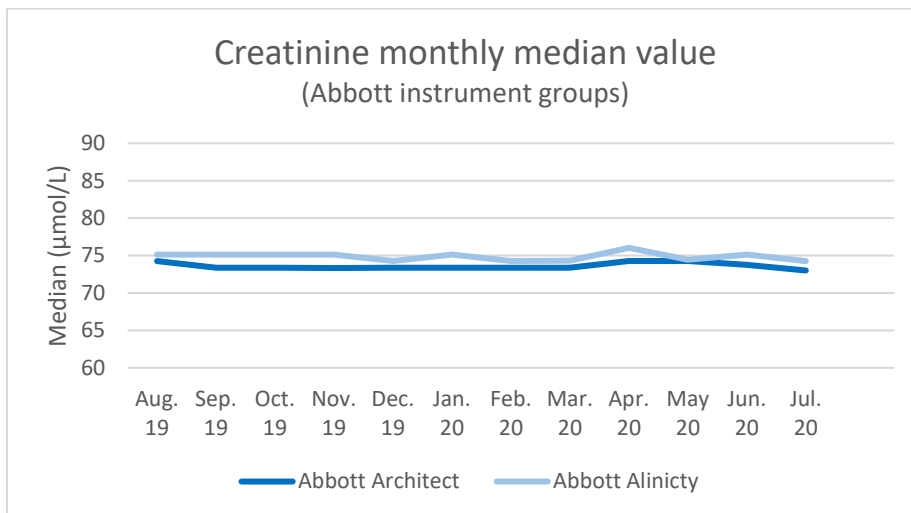
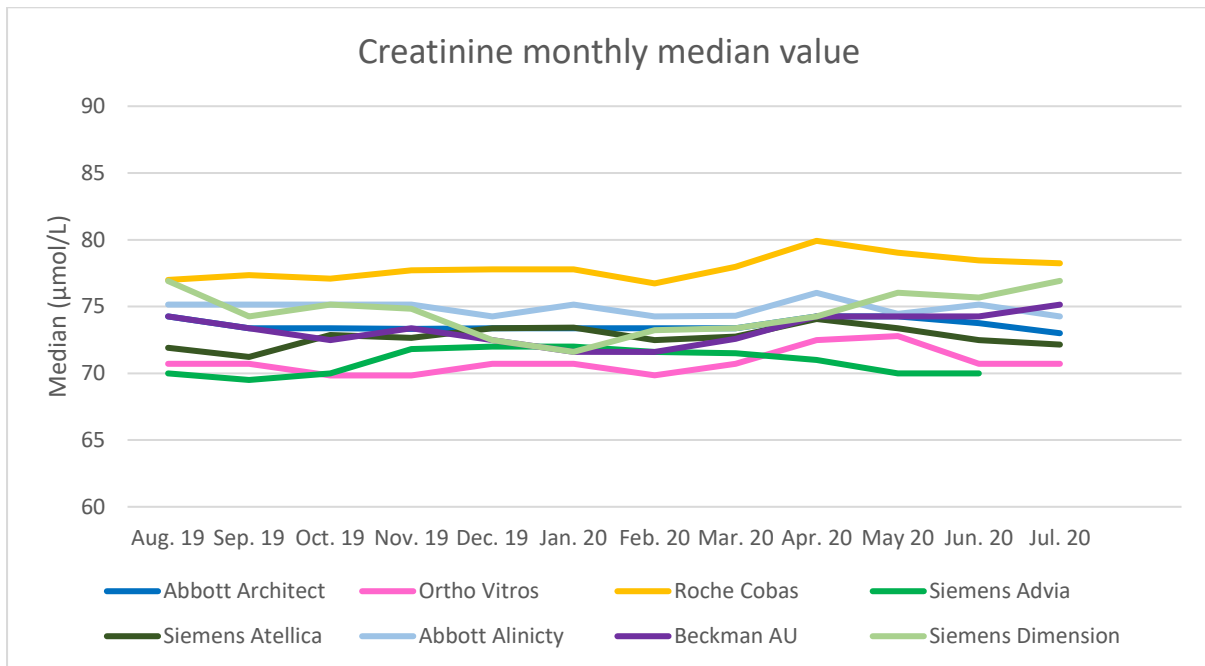
Comments:

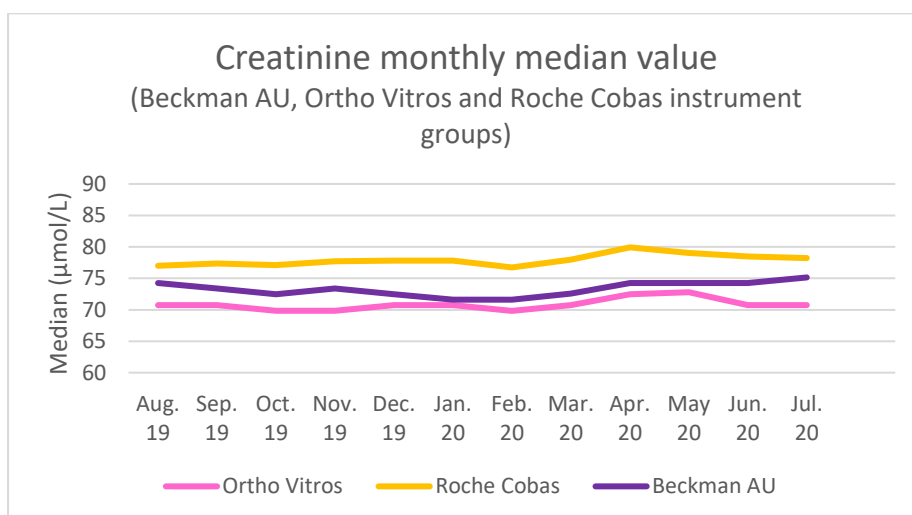
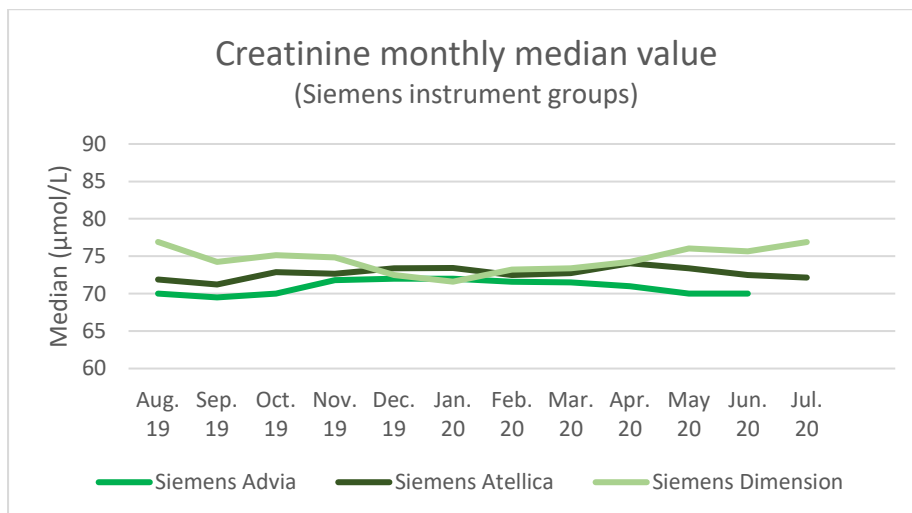
Abbott Architect instruments report the highest median values, and this is also seen in EQA-results from Labquality. We do not know if this is due to an analytical or preanalytical issue. We currently have few Alinity participants, resulting in higher uncertainty.

The difference in median values for cholesterol is clinically relevant if medical doctors use the same decision limits to decide whether the patients should be treated with statins.

Creatinine

Percentiler results:





Instrument group	Creatinine - Median of medians (min-max)	N reported medians (pr. month)
Abbott Alinity	75.1 (74.1 – 76.0)	134
Abbott Architect	73.4 (73.0 – 74.3)	813
Beckman AU	73.4 (71.6 – 75.1)	165
Ortho Vitros	70.7 (69.8 – 72.8)	349
Roche Cobas	77.8 (76.7 – 79.9)	4806
Siemens Atellica	72.7 (71.2 – 74.1)	227
Siemens Advia	71.0 (69.5 – 72.0)	135
Siemens Dimension	74.5 (71.6 – 76.9)	110

Flagger results:

Creatinine	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
	Instrument group	Results above the upper reference limit	
Abbott Alinity	17.0 (12.0-24.0)	1.0 (0.0-2.0)	83
Abbott Architect	10.8 (7.7-16.0)	1.4 (0.5-3.0)	524
Beckman AU	14.0 (10.1-20.0)	4.0 (1.6-7.0)	140
Ortho Vitros	18.0 (11.0-29.0)	6.0 (0.0-12.0)	65
Roche Cobas	12.2 (10.1-15.8)	4.0 (2.6-5.7)	3582
Siemens Atellica	14.0 (11.0-17.0)	8.0 (5.0-11.0)	157
Siemens Advia	8.1 (7.1-9.2)	2.4 (1.9-2.9)	81
Siemens Dimension	17.0 (14.0-22.0)	9.0 (6.0-13.0)	65

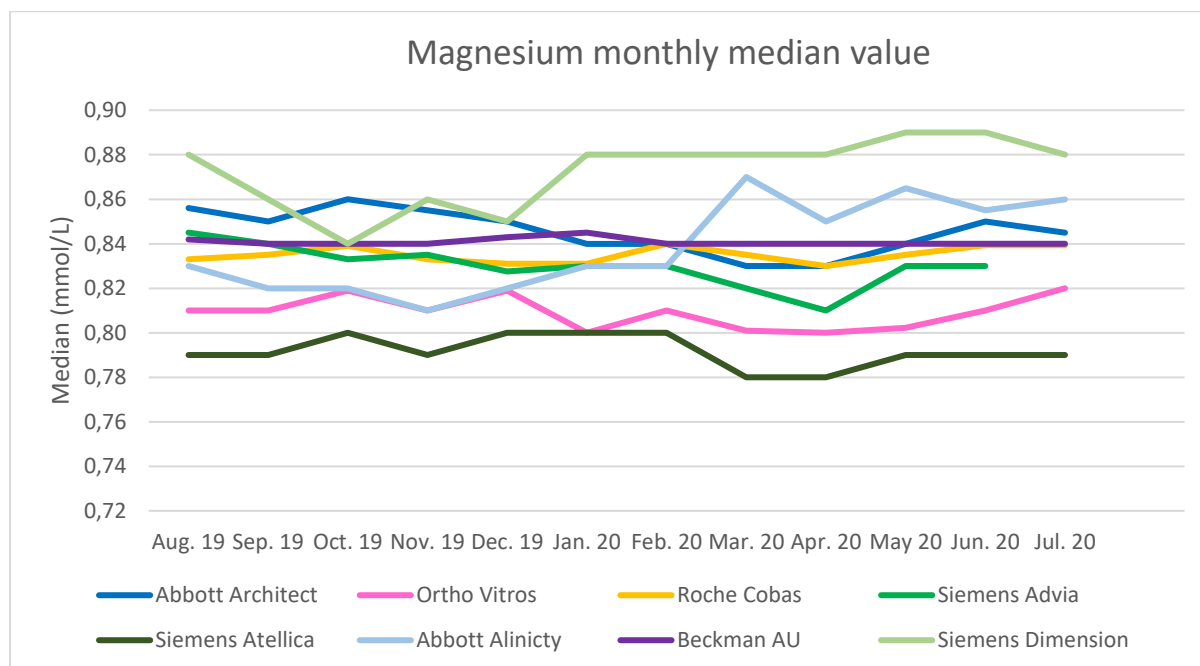
Comments:

The Percentiler results reflect what is already known - that Roche Cobas instruments report the highest median values. A study from 2018 concluded that the CREP2 method on the Cobas c702 overestimates creatinine by 4-6 $\mu\text{mol/L}$ (8). Method differences are also confirmed in a recent study using creatinine as an example for aggregation of commutable EQA results to evaluate metrological traceability and agreement between results (9).

Many laboratories calculate and report eGFR based on the creatinine value, but differences between methods and instruments still cause substantial bias in the overall results (10).

Magnesium

Percentiler results:



Instrument group	Mg - Median of medians (min-max)	N reported medians
Abbott Alinity	0.83 (0.81 – 0.87)	109
Abbott Architect	0.85 (0.83 – 0.86)	599
Beckman AU	0.84 (0.84 – 0.85)	141
Ortho Vitros	0.81 (0.80 – 0.82)	323
Roche Cobas	0.84 (0.83 – 0.84)	4492
Siemens Atellica	0.79 (0.78 – 0.80)	203
Siemens Advia	0.83 (0.81 – 0.85)	114
Siemens Dimension	0.88 (0.84 – 0.89)	101

Flagger results:

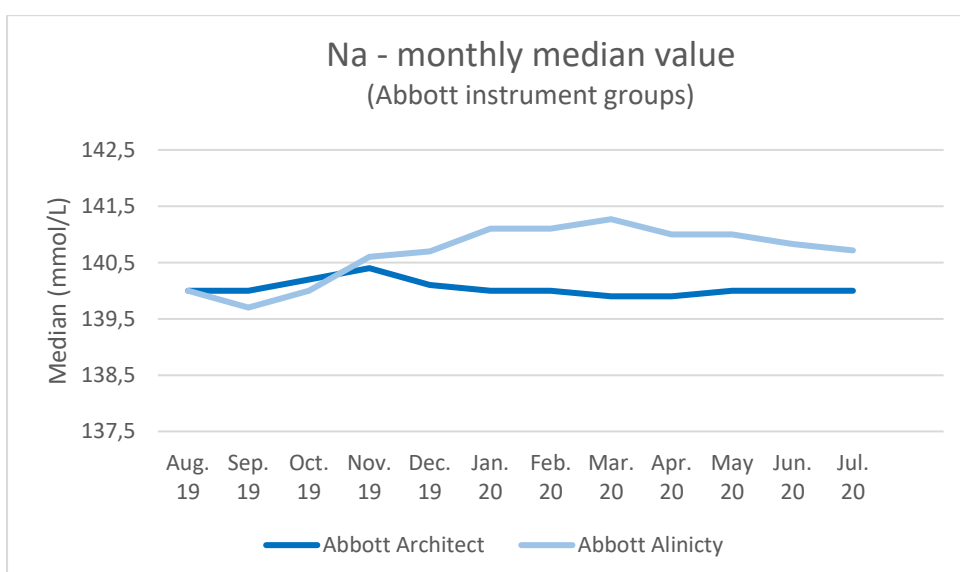
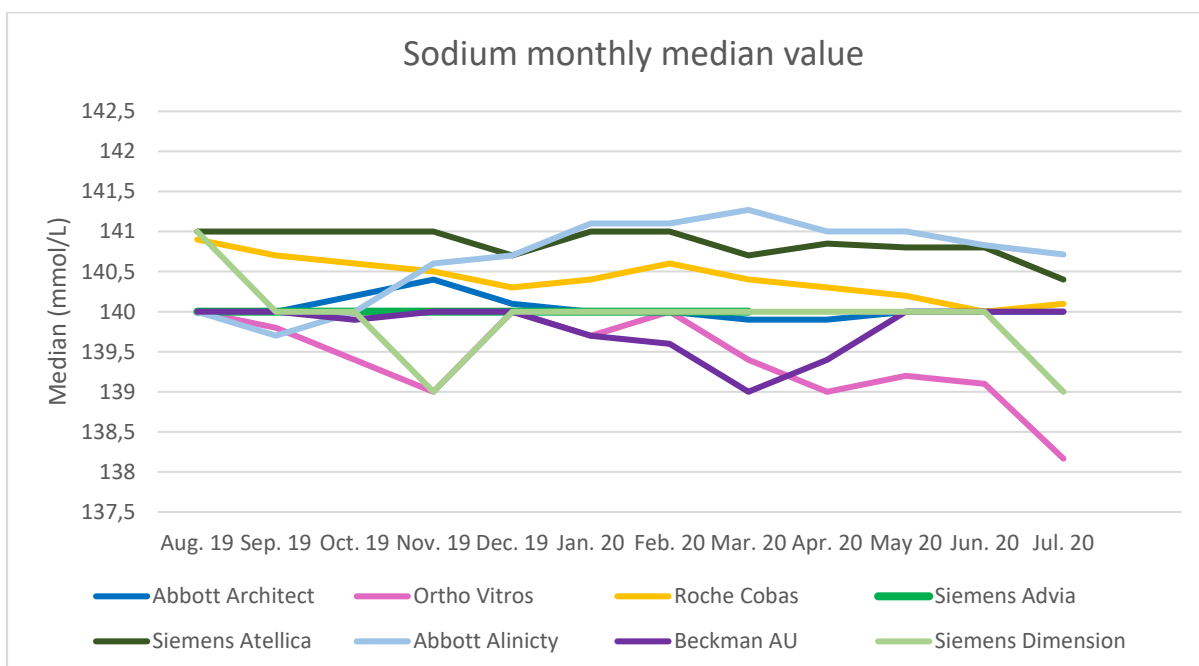
Mg	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
Instrument group	Results above the upper reference limit	Results below the lower reference limit	
Abbott Alinity	0.0 (0.0-2.0)	7.0 (2.0-15.0)	71
Abbott Architect	1.4 (0.0-7.6)	0.0 (0.0-60.)	428
Beckman AU	0.0 (0.0-0.0)	7.0 (0.0-15.0)	122
Ortho Vitros	0.0 (0-0.0)	0.0 (0.0-17.0)	62
Roche Cobas	3.6 (0.0-6.5)	6.7 (3.2-10.7)	3441
Siemens Atellica	0.0 (0.0-0.0)	10.0 (0.0-18.0)	154
Siemens Advia	0.0 (0.0-0.0)	15.4 (7.7-25.0)	80
Siemens Dimension	4.0 (0.0-11.0)	5.0 (0.0-11.0)	64

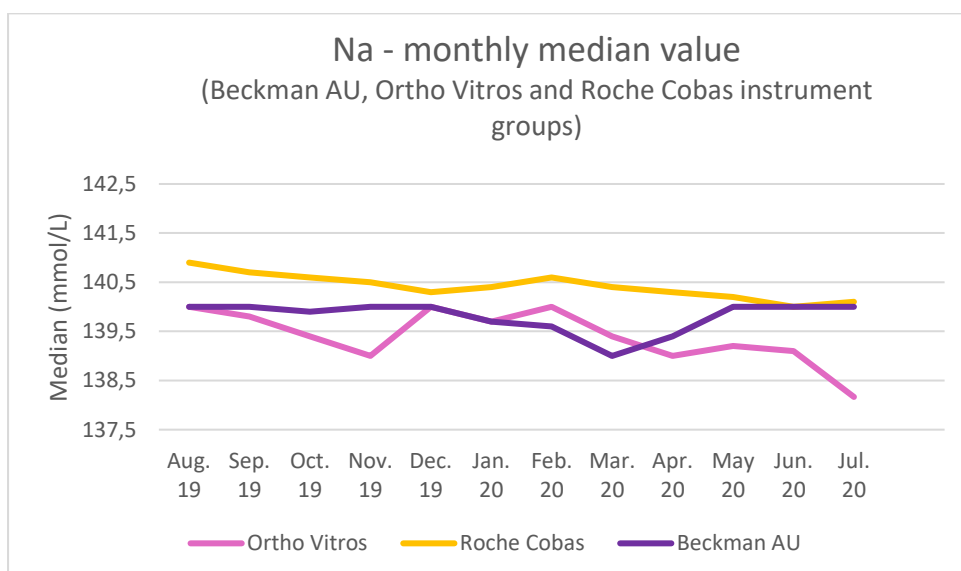
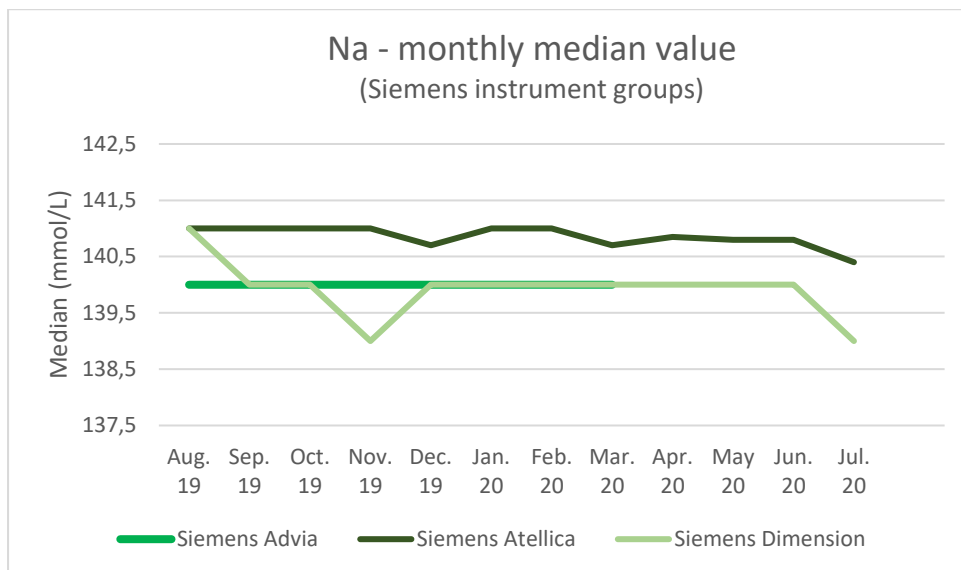
Comments:

We do not find the same method differences for Magnesium in EQA results from Labquality. A possible explanation is that the differences may be caused by preanalytical conditions, and it has been suggested that sample tubes may be influential.

Sodium (Na)

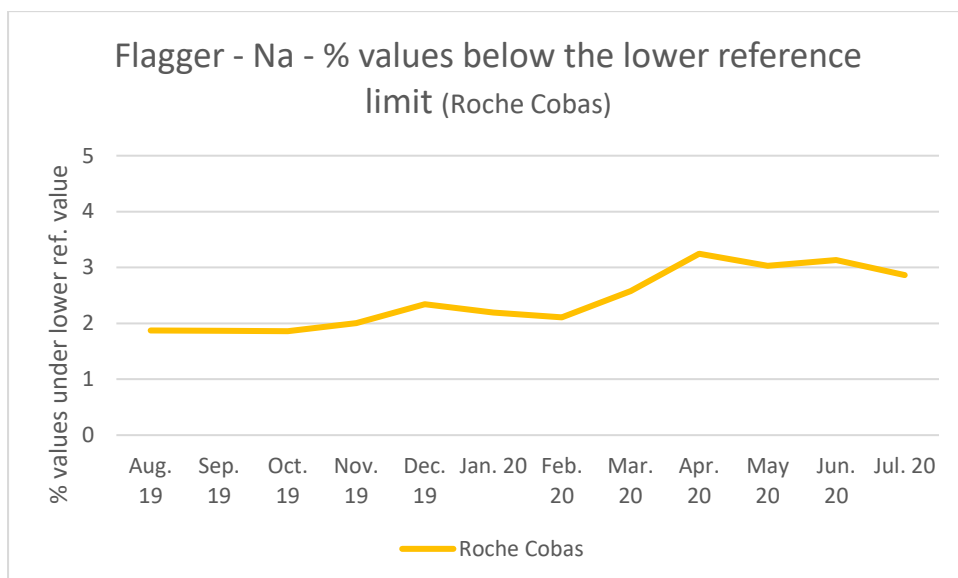
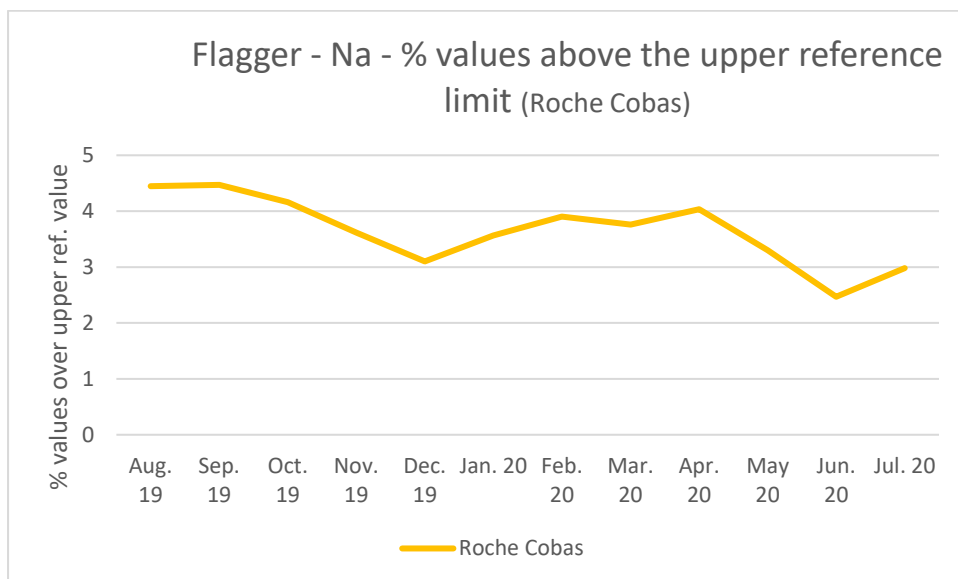
Percentiler results:





Instrument group	Na - Median of medians (min-max)	N reported medians (pr. month)
Abbott Alinity	140.8 (139.7 – 141.3)	141
Abbott Architect	140.0 (139.9 – 140.4)	720
Beckman AU	140.0 (139.0 – 140.0)	155
Ortho Vitros	139.4 (138.2 – 140.0)	340
Roche Cobas	140.4 (140.0 – 140.9)	4527
Siemens Atellica	140.9 (140.4 – 141.0)	173
Siemens Advia	140.0 (140.0 – 140.0)	142
Siemens Dimension	140.0 (139.0 – 141.0)	141

Flagger results:



Results above the upper reference limit

Na	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
	Results above the upper reference limit	Results below the lower reference limit	
Abbott Alinity	0.0 (0.0-2.0)	4.0 (2.0-6.0)	115
Abbott Architect	0.0 (0.0-1.4)	4.8 (3-9.0)	478
Beckman AU	0.0 (0.0-0.4)	6.0 (3.0-11.0)	132
Ortho Vitros	0.0 (0.0-0.0)	9.0 (0.0-16.0)	65
Roche Cobas	3.5 (1.4-7.0)	2.4 (1.6-3.9)	3163
Siemens Atellica	2.0 (1.0-4.0)	7.0 (5.0-9.0)	103
Siemens Advia	0.0 (0.0-0.1)	3.5 (2.9-4.1)	81
Siemens Dimension	0.0 (0.0-1.0)	3.0 (1.0-6.0)	65

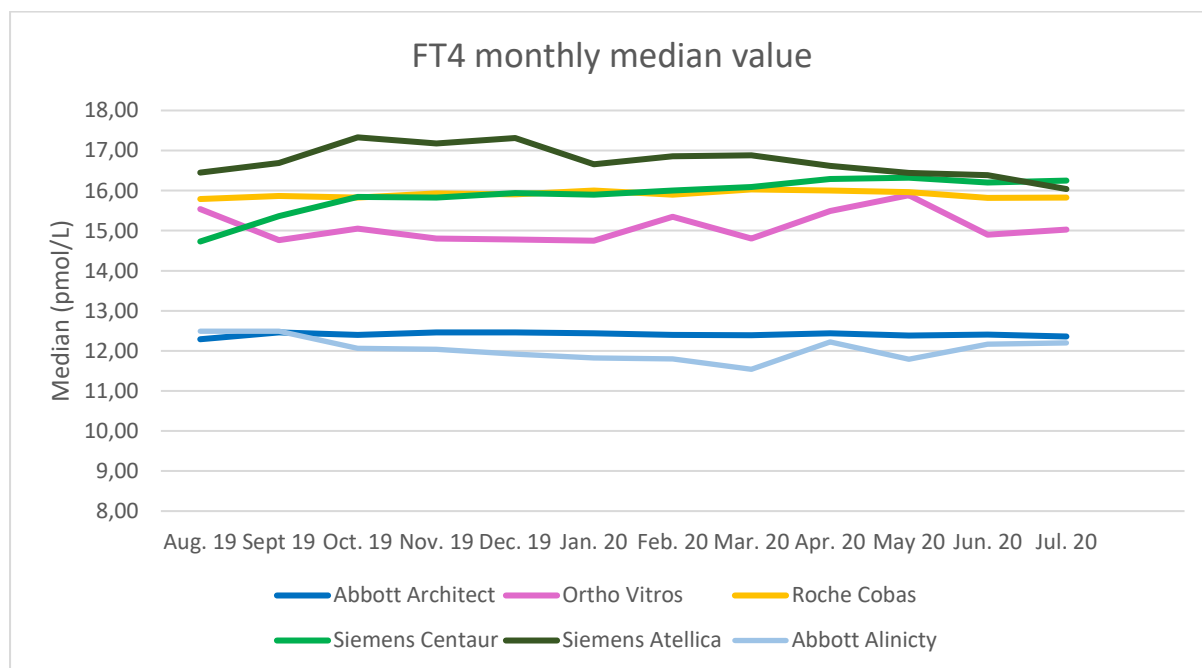
Comments:

There are two major method groups for sodium; ISE direct and ISE indirect methods. When ISE indirectly measured sodium is compared to an ISE directly measured sodium method, there is a systematic difference (11).

Overall, all methods are in reasonable agreement. To our knowledge, there is no published explanation for the changes in percent flagged results for the large Cobas group.

FT4

Percentiler results:



Instrument group	FT4 - Median of medians (pmol/L) (min-max)	N reported medians (pr. month)
Abbott Alinity	12.1 (11.5 – 12.5)	104
Abbott Architect	12.4. (12.3 – 12.5)	577
Ortho Vitros	15.0 (14.8 – 15.9)	100
Roche Cobas	15.9 (15.8 – 16.0)	1704
Siemens Atellica	16.7 (16.0 – 17.3)	104
Siemens Centaur	16.0 (14.7 – 16.3)	289

Flagger results:

FT4 Instrument group	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
	Results above the upper reference limit	Results below the lower reference limit	
Abbott Alinity	1.0 (0.0-2.1)	2.0 (0.0-3.4)	77
Abbott Architect	1.3 (0.0-2.3)	1.9 (0.1-3.2)	390
Ortho Vitros	0.0 (0.0-2.0)	0.0 (0.0-7.0)	42
Roche Cobas	5.0 (0.0-10.0)	2.0 (0.0-7.0)	638
Siemens Atellica	6.0 (3.0-9.0)	1.0 (0.0-4.0)	94
Siemens Centaur	4.2 (2.5-6-3)	1.3 (2.1-2.3)	144

Comments:

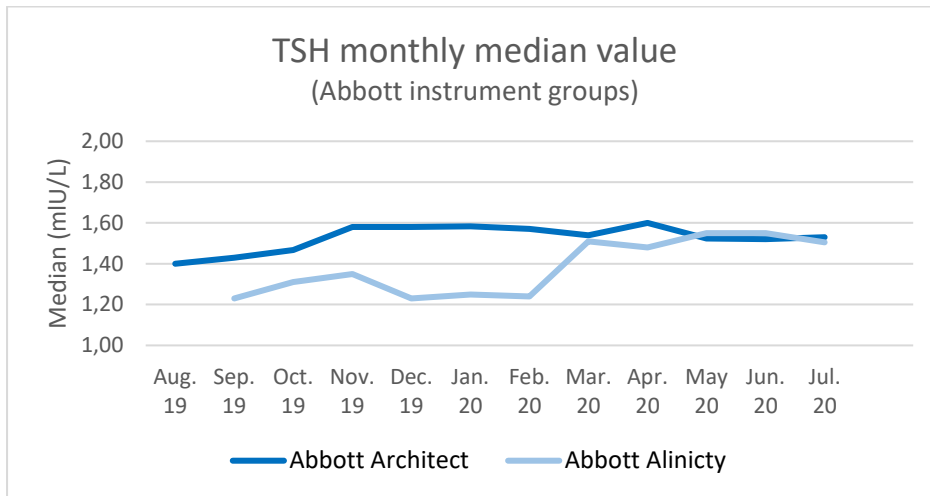
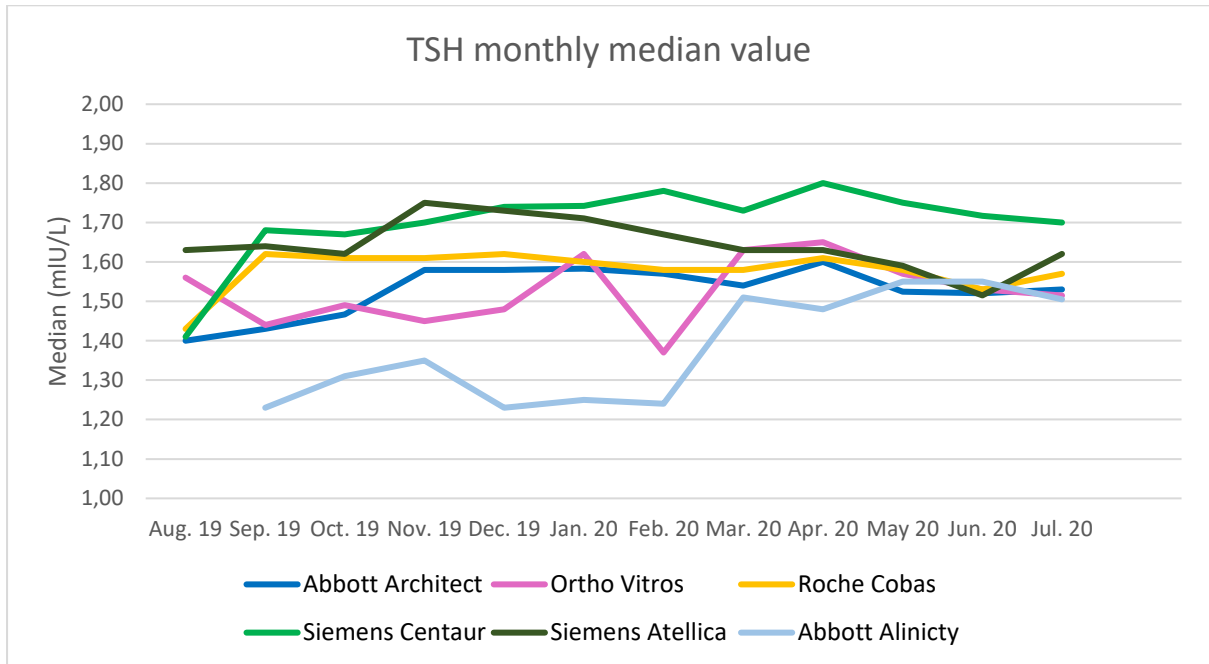
Both Abbott instrument groups report median values at a lower level than all other instrument groups.

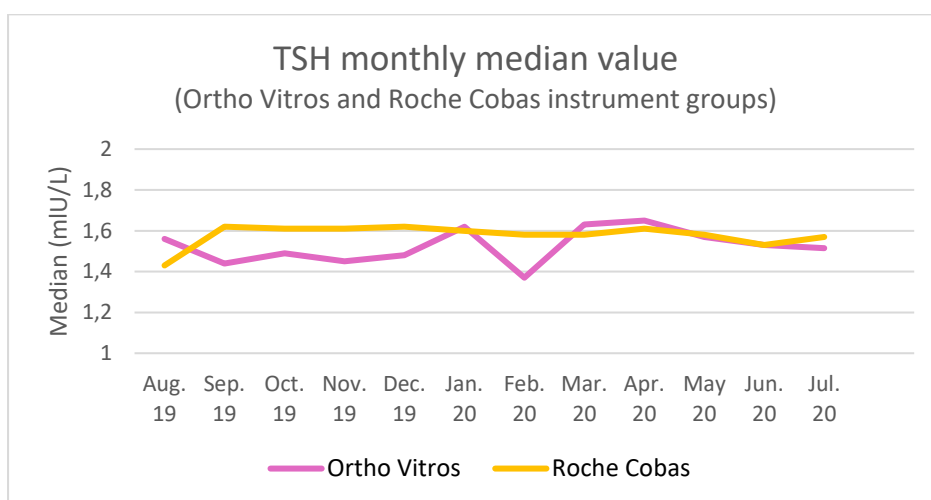
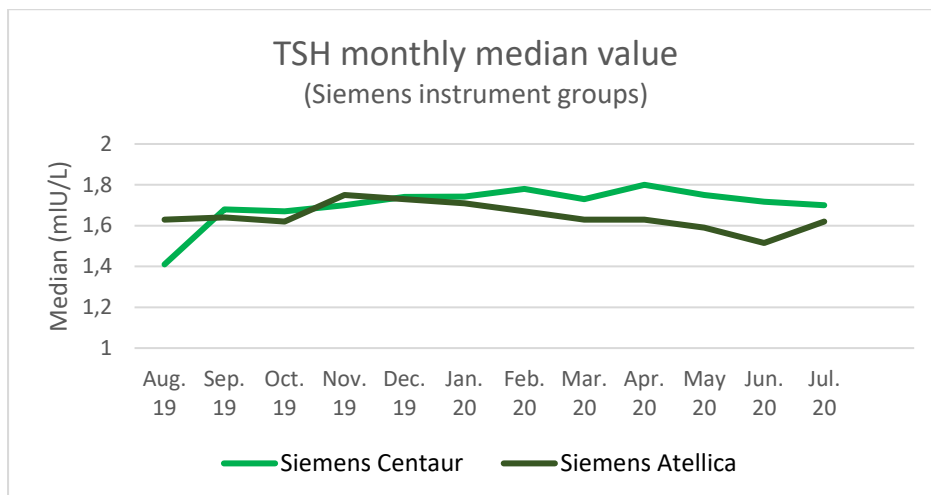
There is a need for harmonization, which is the aim of the IFCC Committee for Standardization of Thyroid Function Tests. This Committee indeed made both a reference method and a reference material for standardization of free Thyroxine (FT4) immunoassays available (12), so that theoretically, one should expect the instruments to measure at the same level.

Clinical practice uses method-independent cut-off values, which makes harmonization particularly important and any differences clinically relevant for patient treatment.

TSH

Percentiler results:





Instrument group	TSH - Median of medians (min-max)	N reported medians
Abbott Alinity	1.35 (1.23 – 1.55)	138
Abbott Architect	1.54 (1.40 – 1.60)	603
Ortho Vitros	1.52 (1.37 – 1.65)	100
Roche Cobas	1.59 (1.43 - 1.62)	2033
Siemens Atellica	1.63 (1.52 – 1.75)	139
Siemens Centaur	1.72 (1.41 – 1.80)	301

Flagger results:

TSH	Median value (%) (25-percentile-75 percentile)		N reported medians (average pr. month)
	Results above the upper reference limit	Results below the lower reference limit	
Abbott Alinity	4.0 (1.0-11.0)	7.7 (6.0-9.0)	80
Abbott Architect	8.6 (4.7-11.7)	6.7 (4.5-8.9)	394
Ortho Vitros	0.0 (0.0-6.0)	13.0 (8.0-19.0)	43
Roche Cobas	6.0 (3.0-10.0)	5.0 (2.0-8.3)	739
Siemens Atellica	8.0 (4.0-13.0)	12.0 (8.0-16.0)	119
Siemens Centaur	5.0 (3.5-6.7)	5.0 (3.8-7.0)	149

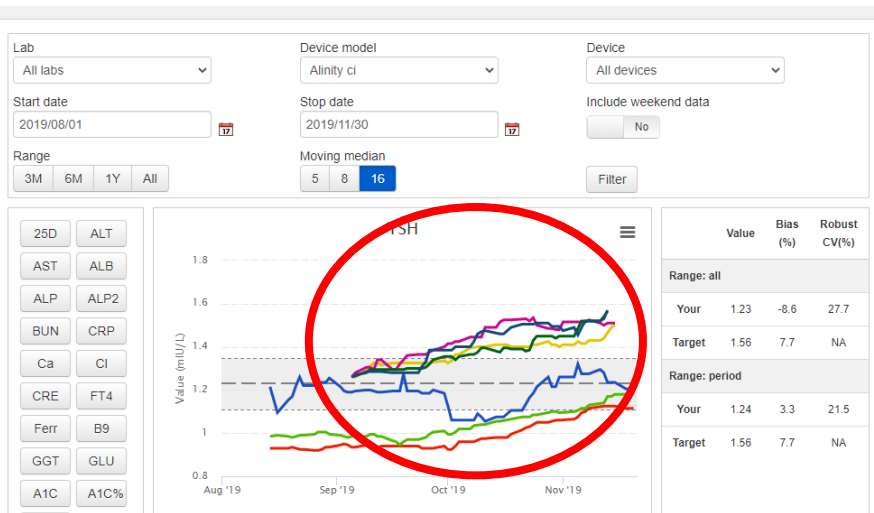
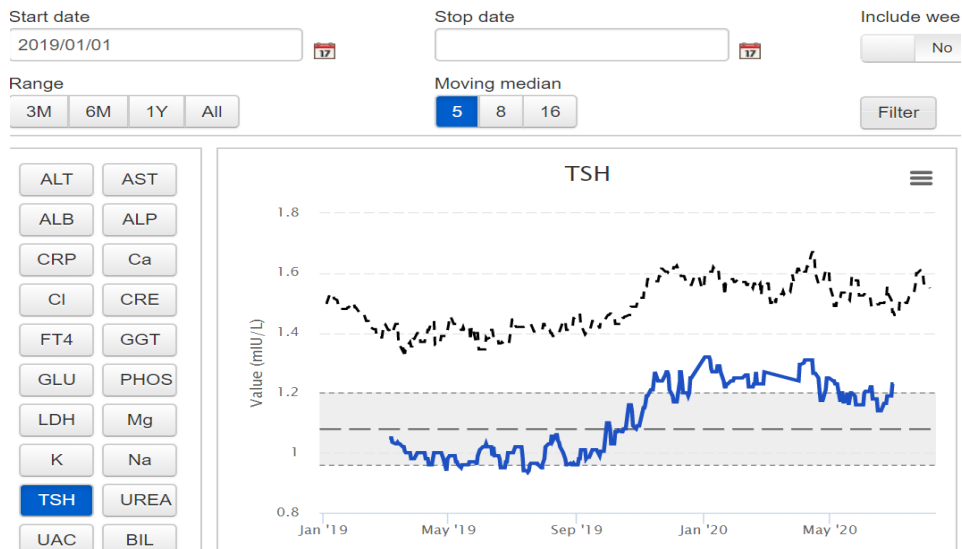
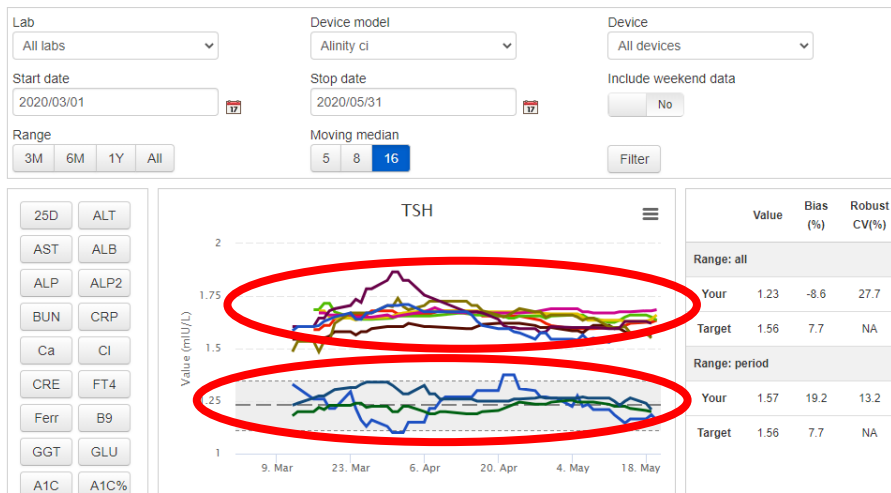
Comments:

There now seems to be a better harmonization of TSH than before, which can most probably be ascribed to the efforts done by the earlier mentioned IFCC Committee for Standardization of Thyroid Function Tests (13). In this publication, it was demonstrated that TSH harmonization by use of a reference material made available by the Committee is feasible without dramatic consequences for the companies (e.g., new submission for approval by authorities like FDA).

In both the Architect and Alinity group there are two subgroups which have their TSH median values at different levels (illustrated for the Alinity group in the figure below by the red circles around the two subgroups/medians). To elucidate whether this points to an analytical bias between the concerned laboratories, one could let them analyse the same panel of native patient samples. If this experiment would give results that agree well between the two subgroups, the existence of a bias would be ruled out. Hence, one would have to accept that there is another cause for the difference in median values. For TSH it might be hypothesized that the difference is due to laboratories testing different patient populations, namely primary (first line testing) versus secondary care patients (dedicated testing for management of thyroid disease). Alternatively, the difference could be caused by preanalytical conditions, like differences in sample tubes.

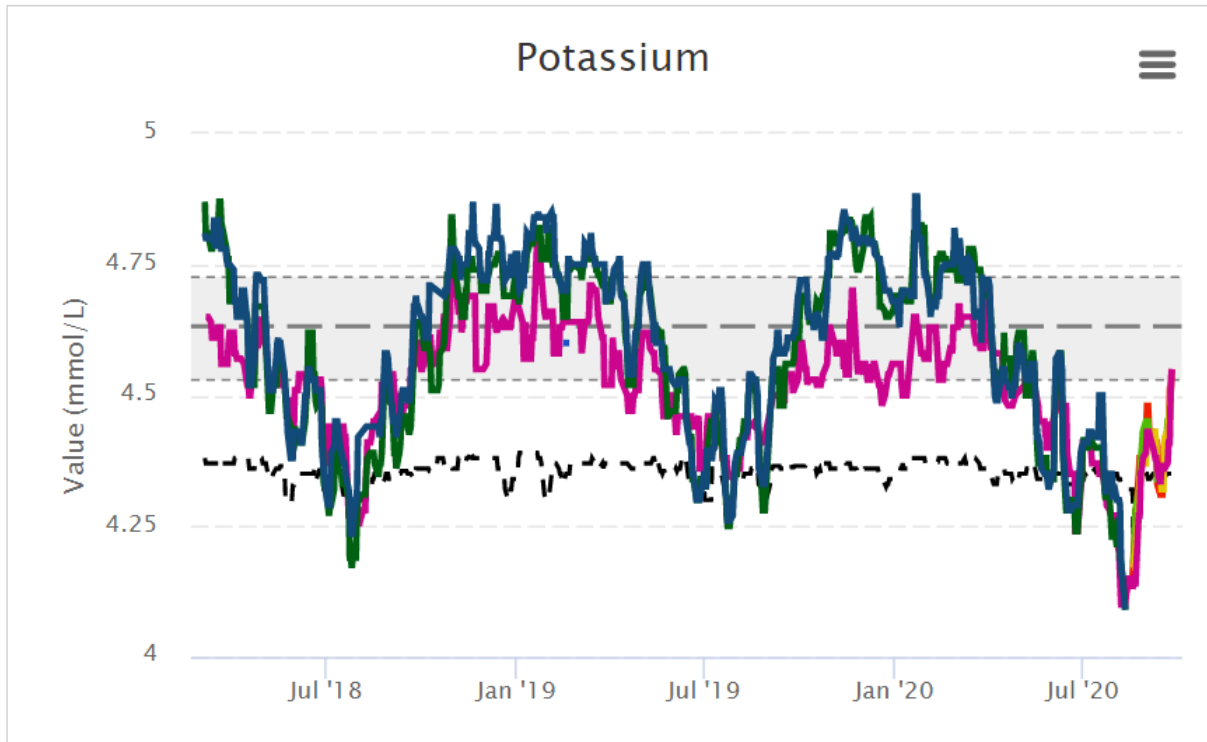
Therefore, until the comparison with the aforementioned panel of native sera is done, we recommend that the Architect and Alinity laboratories restrict the use of the Percentiler to interpretation of their own long-term stability and not for judging whether they are biased in comparison to other laboratories belonging to their peer group.

Many of the Abbott instruments showed an increase in the median value during autumn 2019 (illustrated in the 2 figures below), for which we have no obvious explanation.



Potassium

To illustrate how preanalytical problems can be found in the Percentiler program, we have added a figure for one laboratory from 01/03/2018 to 30/09/2020. The typical profile shows high values during the winter, and lower median values (actually normal) in the summer.

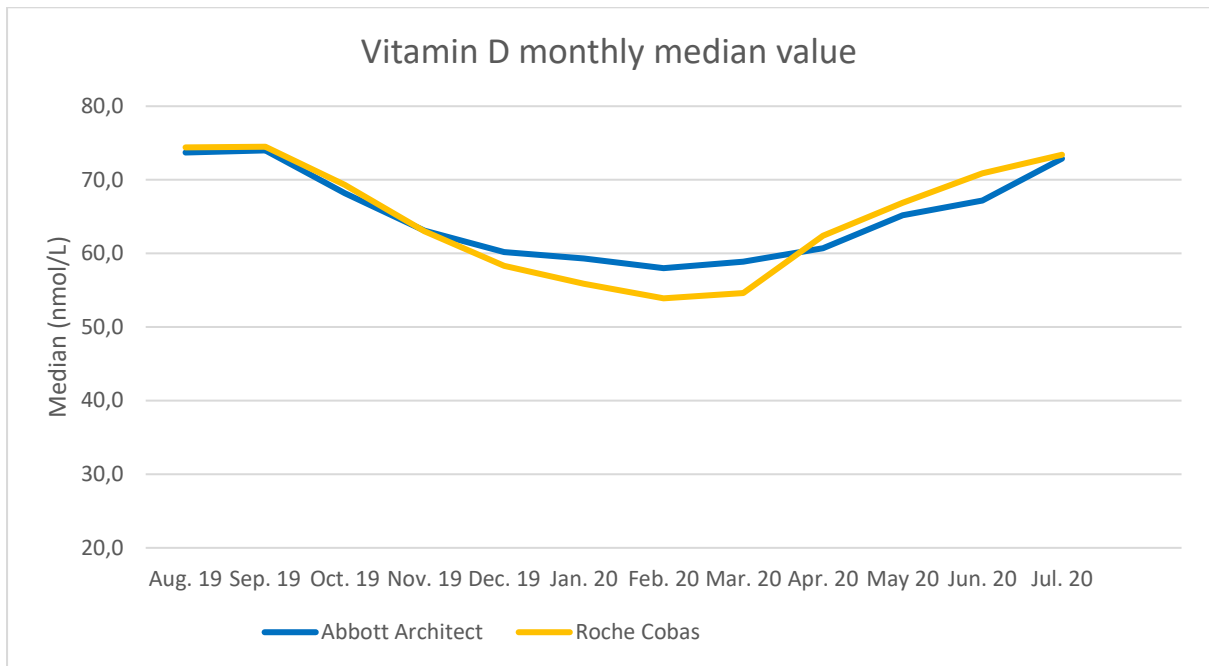


Other results from the Percentiler program

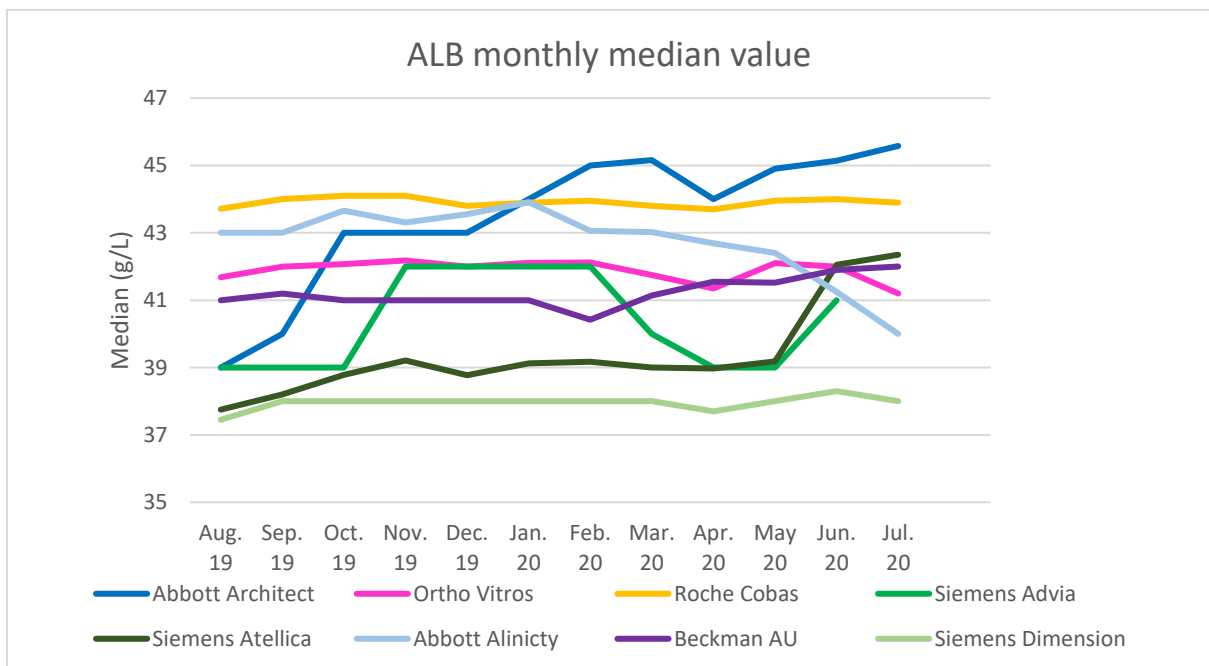
Below, you will find figures for the monthly median values for several other analytes, and for most of the instrument groups. We believe that most of the observed differences are analytical differences between instrument groups and methods. Even so, the results for each analyte must be interpreted in the context of:

- The population the results represent. For instance, in populations where folate is added to enrich the diet, the measured folate level is expected to be higher than in populations where food is not enriched. Other vitamins may also be affected by the local diet.
- Variation throughout the year, as seen for 25-Hydroxy vitamin D.
- Why a test is requested. For instance, HbA1c can be used both to diagnose diabetes and to monitor diabetic patients, and the use varies between countries.
- Preanalytical factors. For some analytes we know that tubes and sample material is important, for example when measuring potassium. For glucose, some laboratories report only fasting glucose values while others report the median value of all outpatient results.
- Analytical factors. Even within instrument groups there may be different methods and reagents in use for the same analyte. Also, some laboratories factorize their results. A good example is albumin, where the Percentiler does not differentiate between Brom Cresol Green (BCG) and Brom Cresol Purple (BCP) methods. BCG methods report higher values than BCP methods, and several of the instrument vendors offer both BCG and BCP methods. In future versions of the Percentiler, we will consider distinguishing between these methods.
- The COVID-19 pandemic. We know that during the pandemic, several participants have analysed fewer samples than usual, and the population may also differ from the usual population.

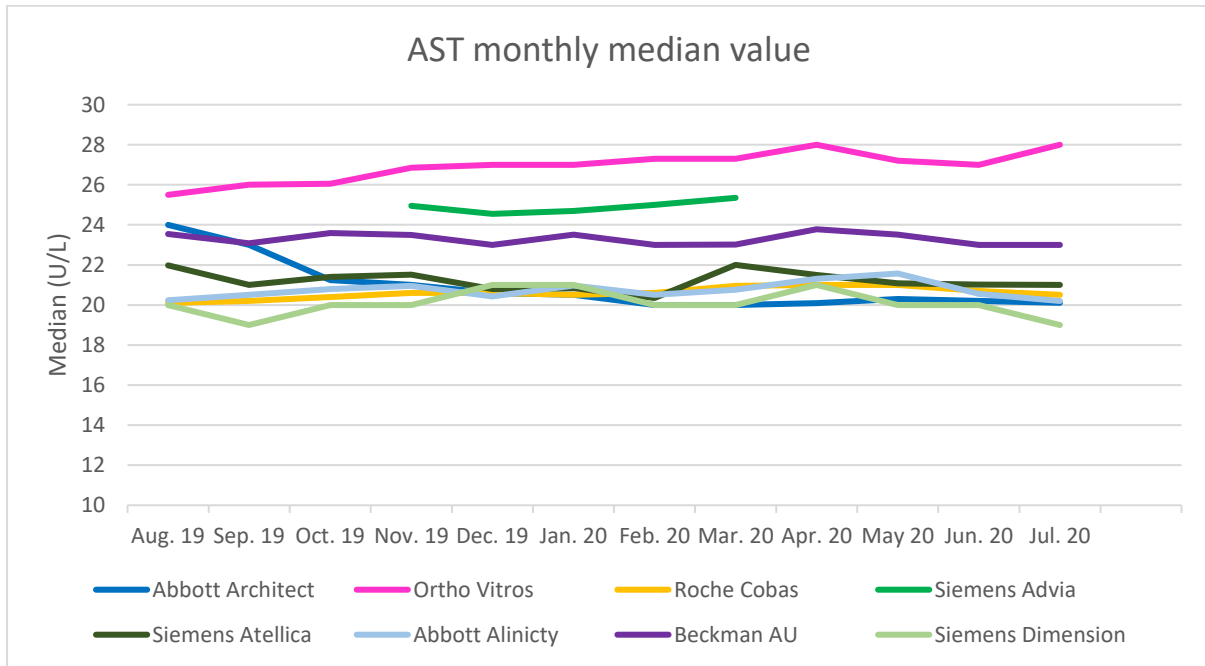
25-Hydroxy vitamin D



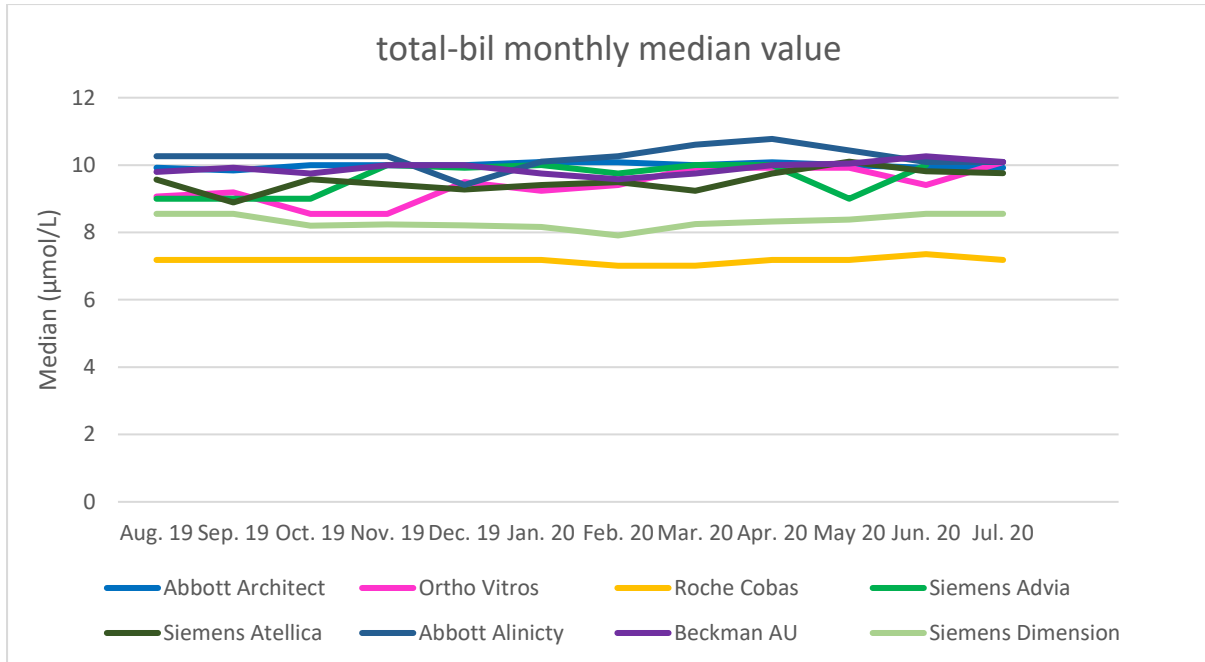
ALB



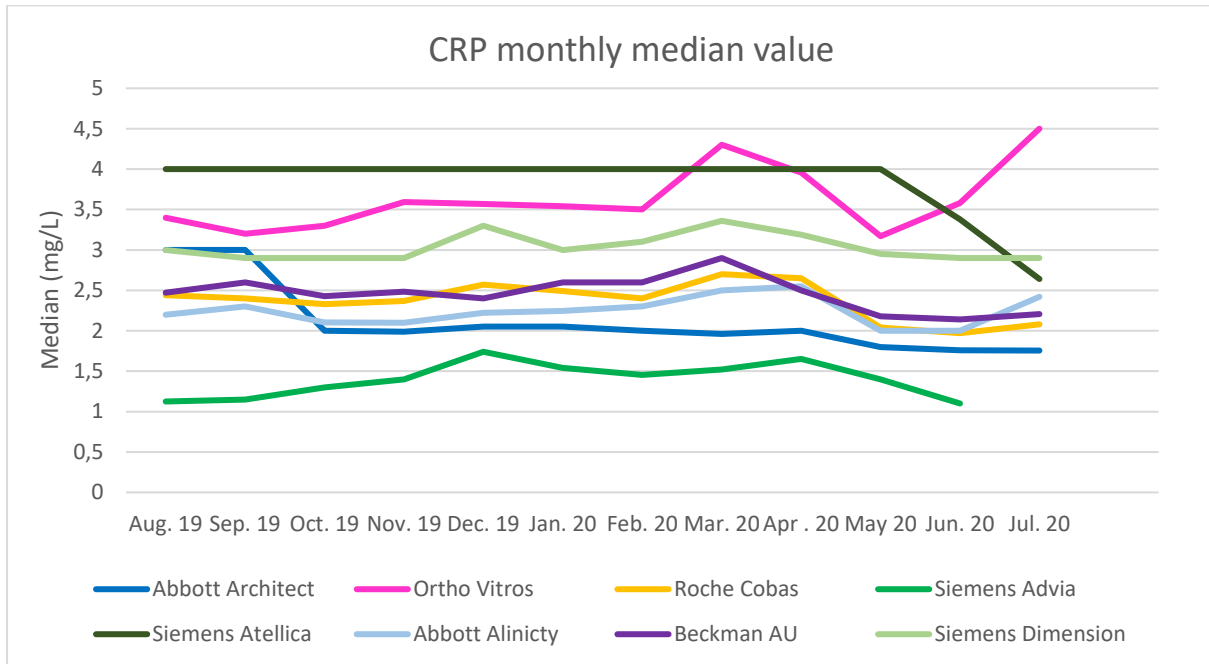
AST



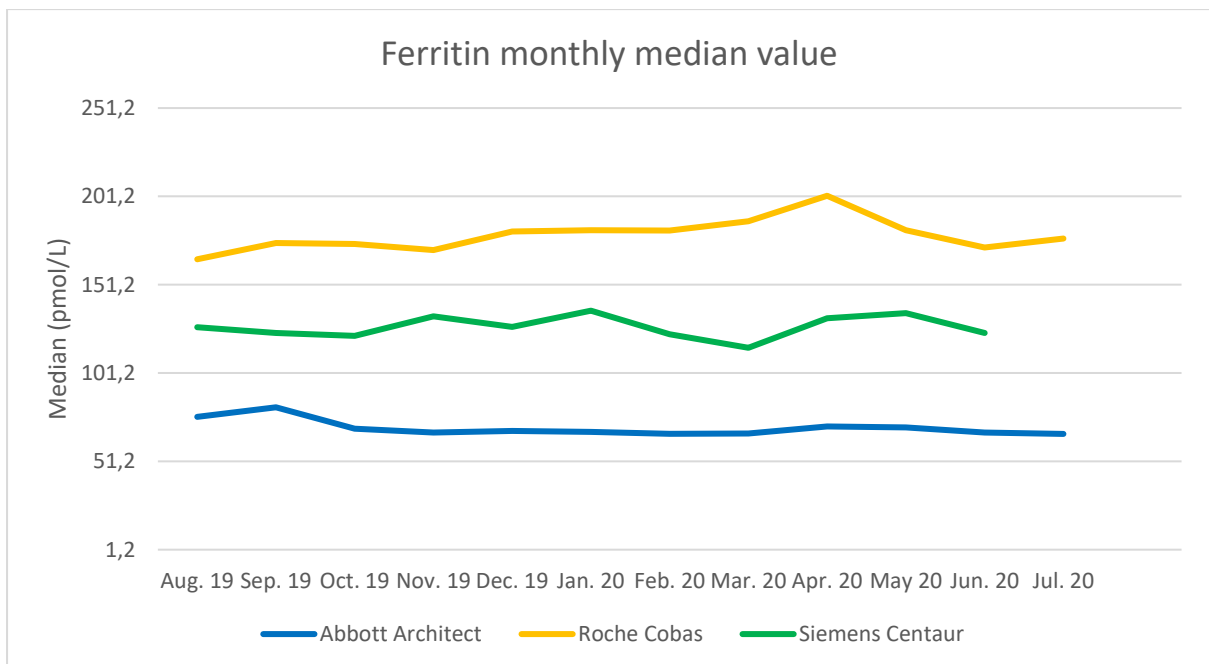
Bilirubin (total)



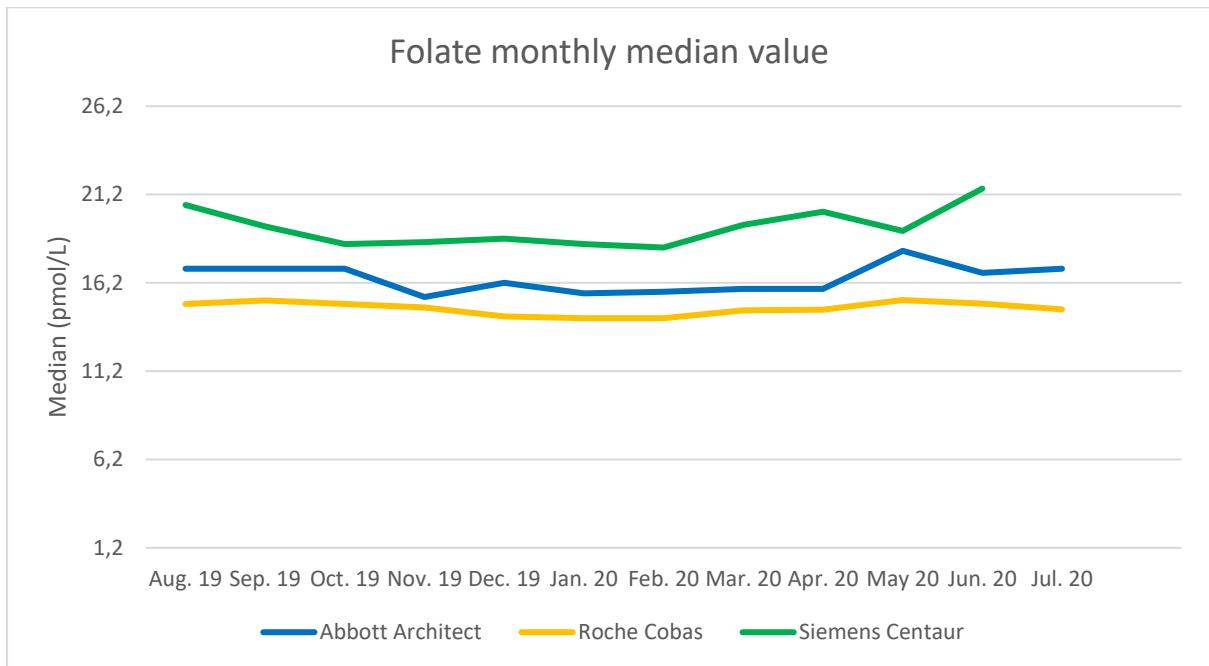
CRP



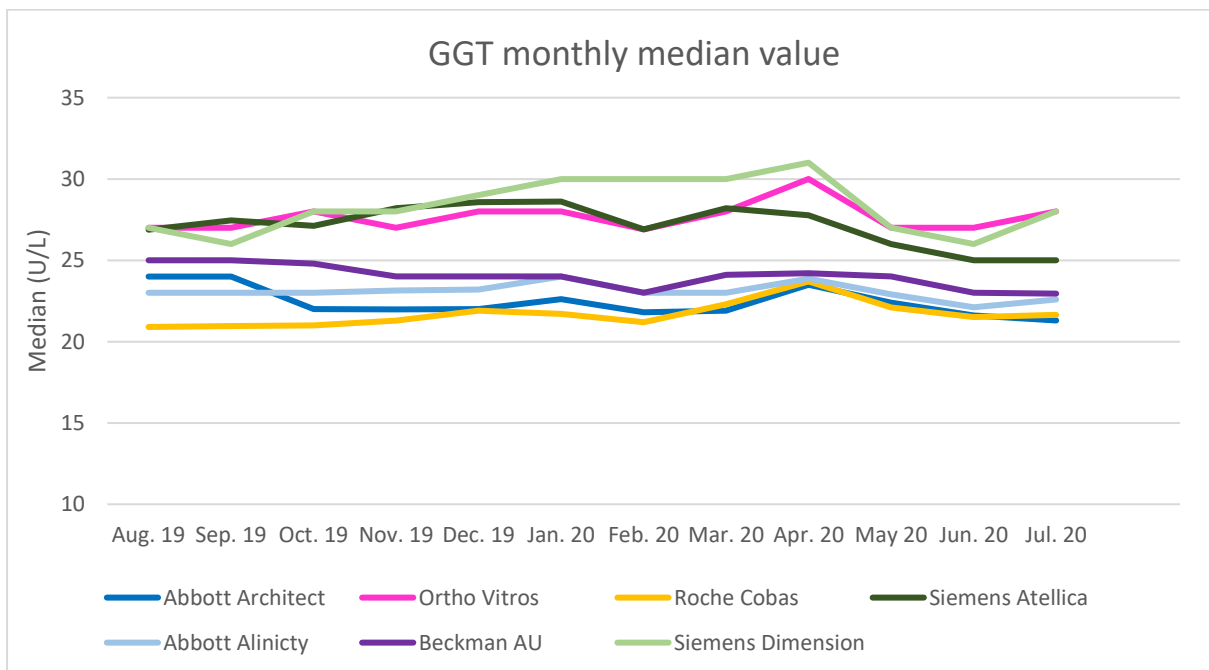
Ferritin



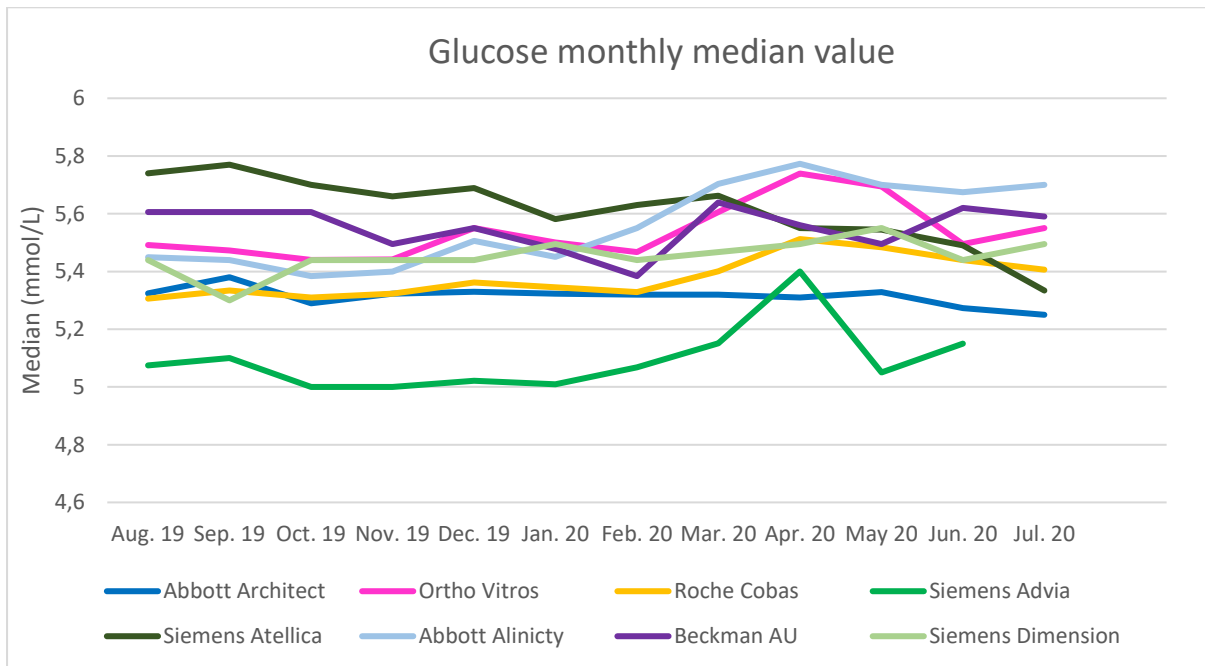
Folate



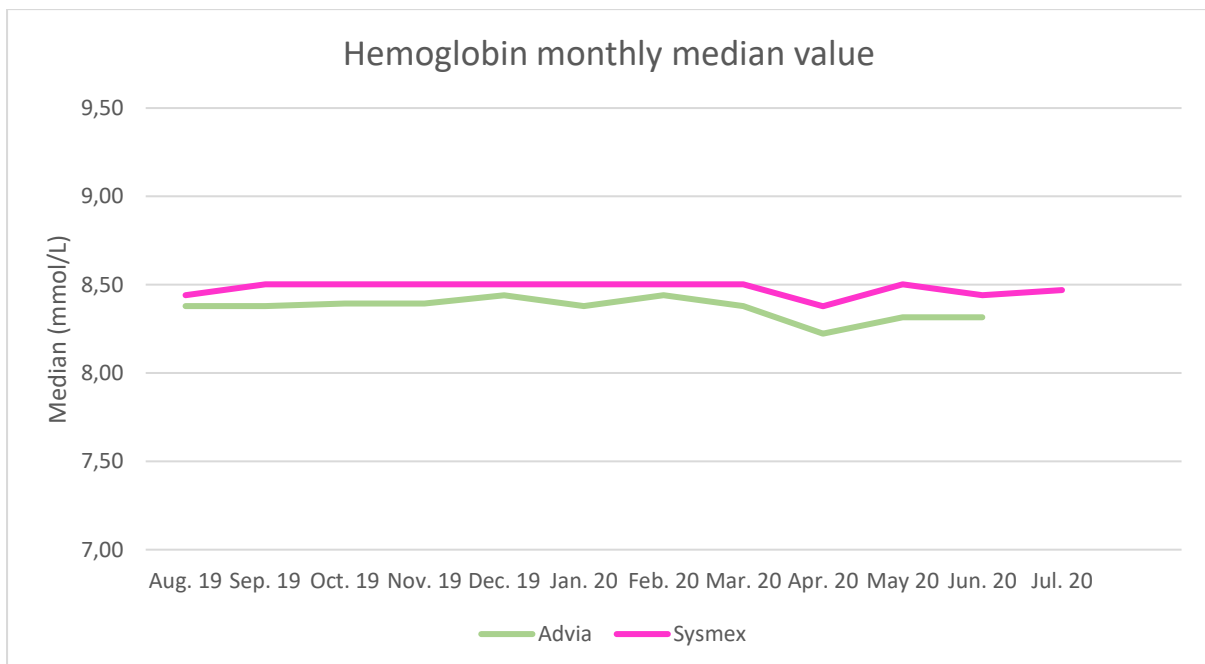
GGT



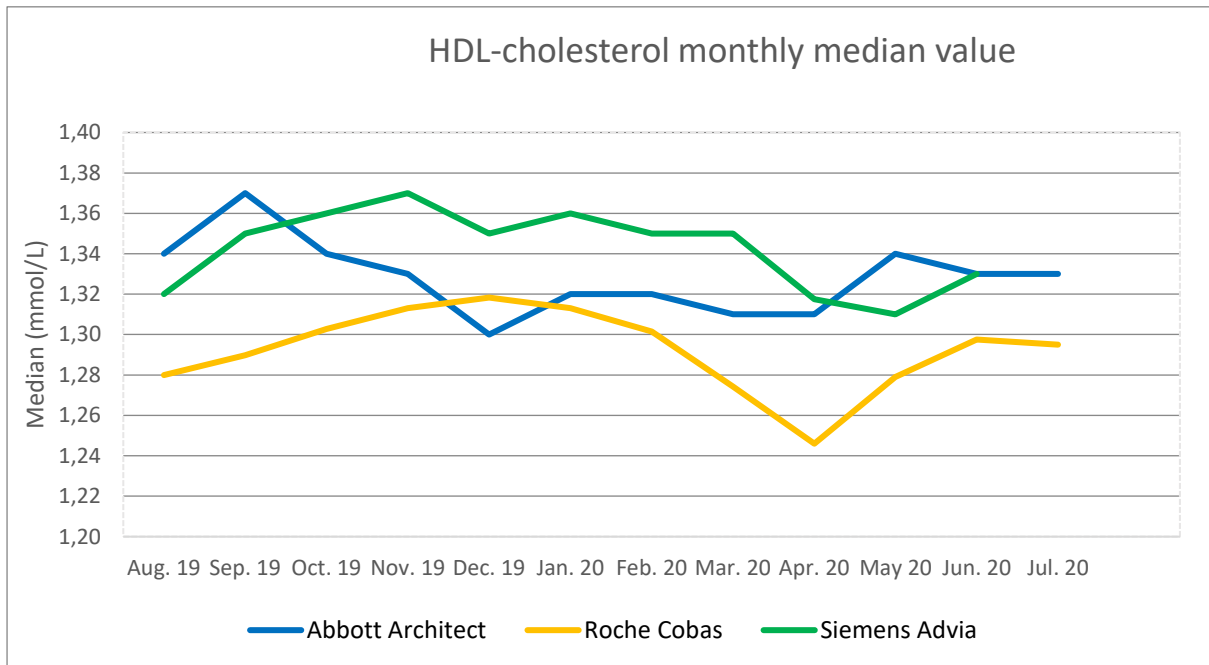
Glucose



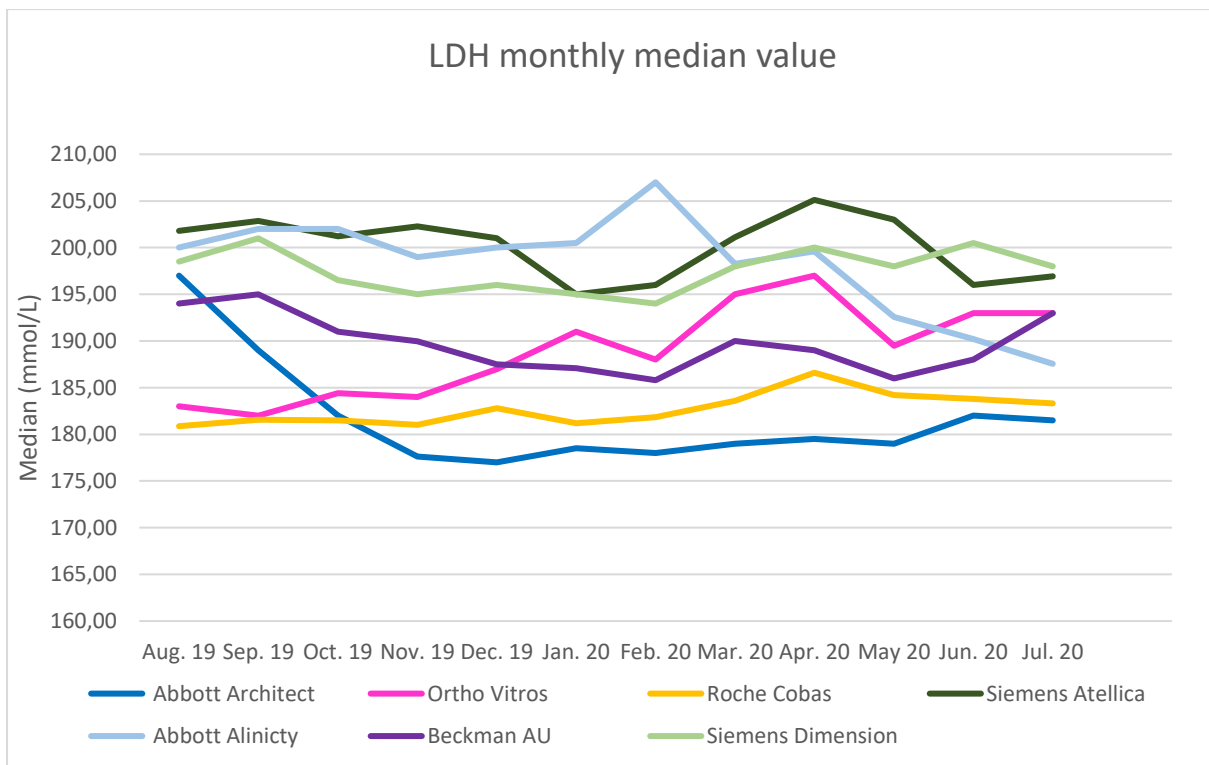
Hb



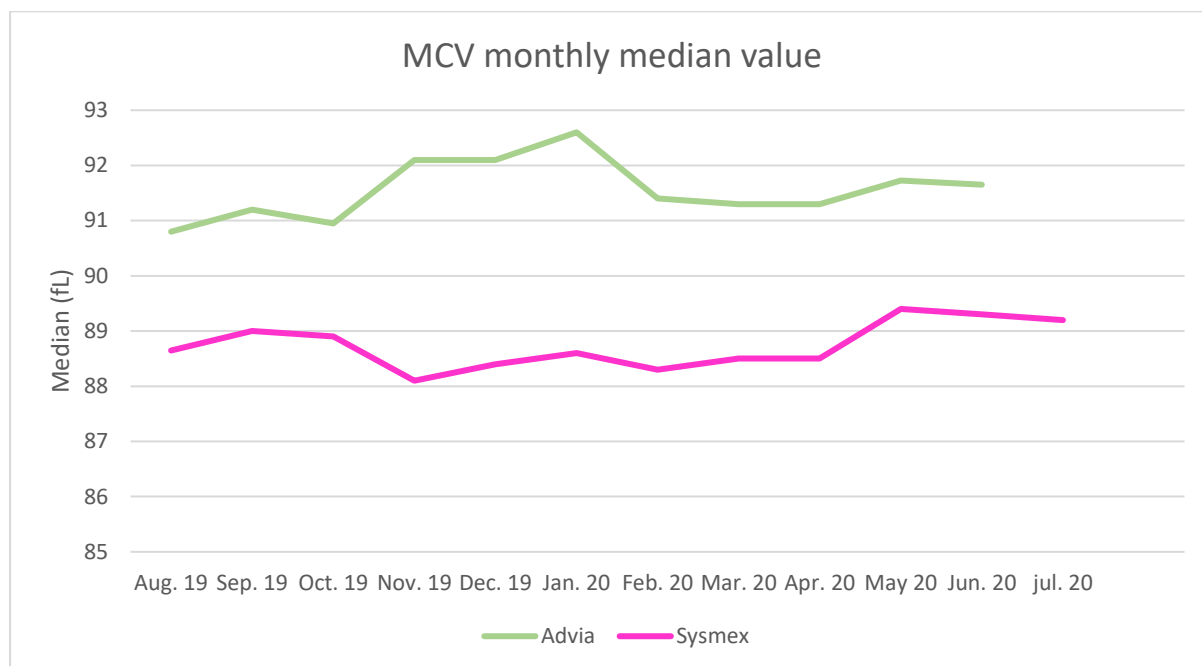
HDL-cholesterol



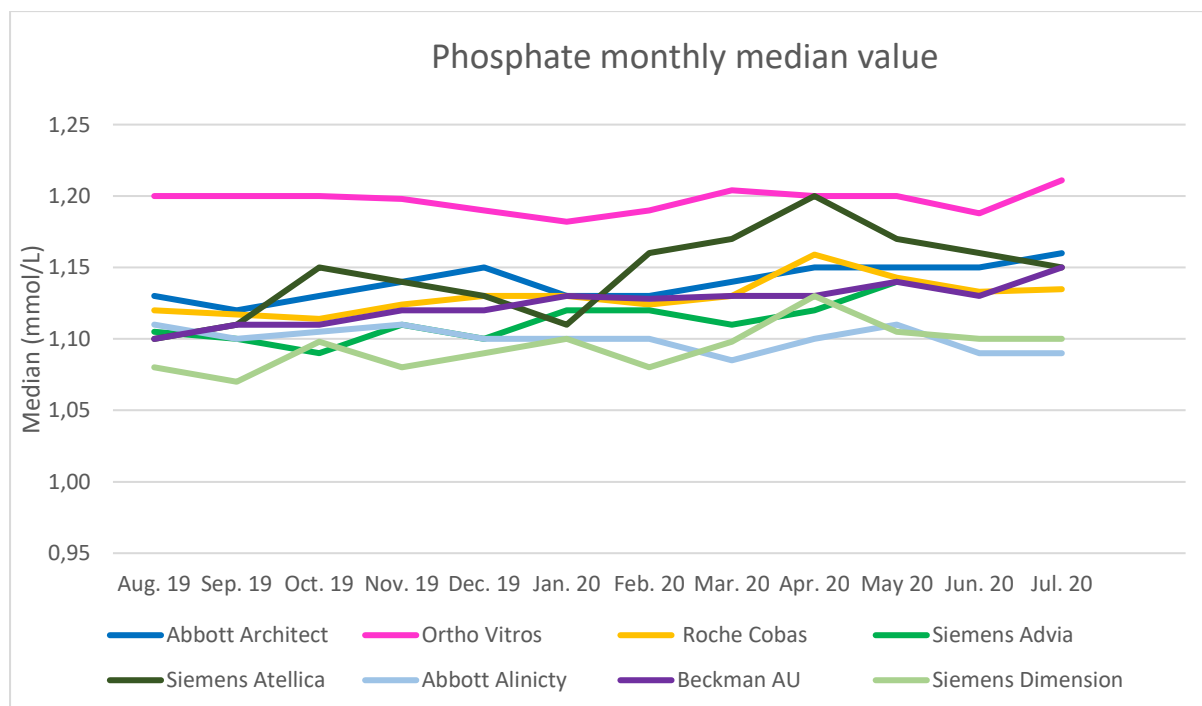
LDH



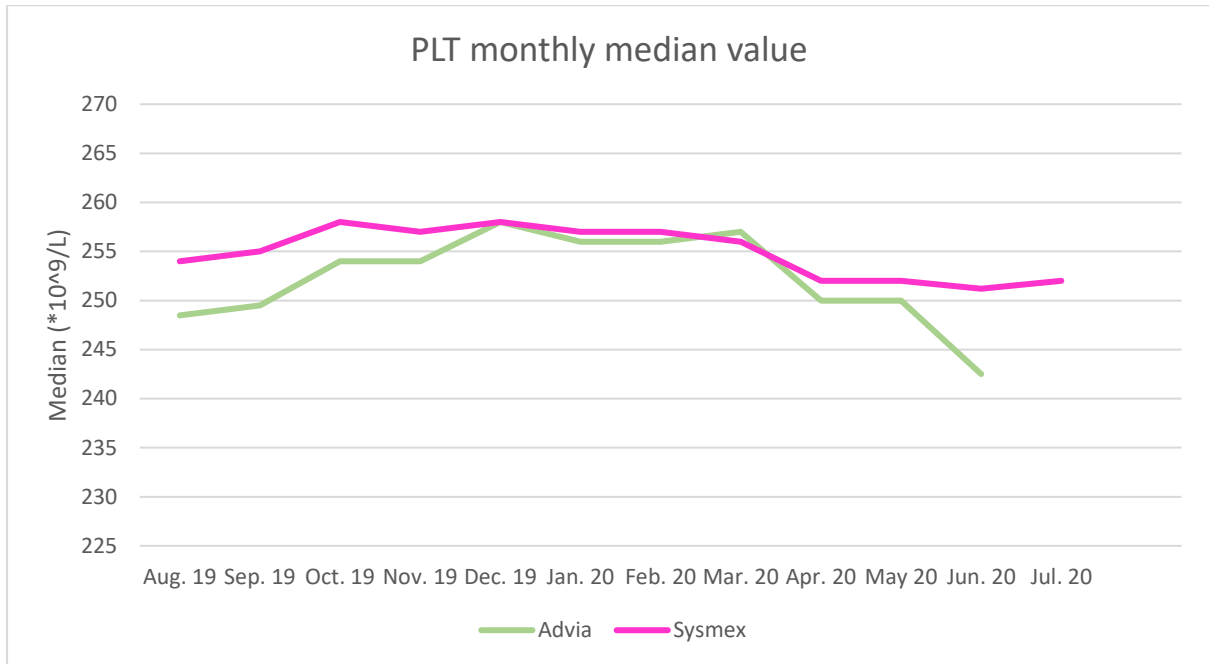
MCV



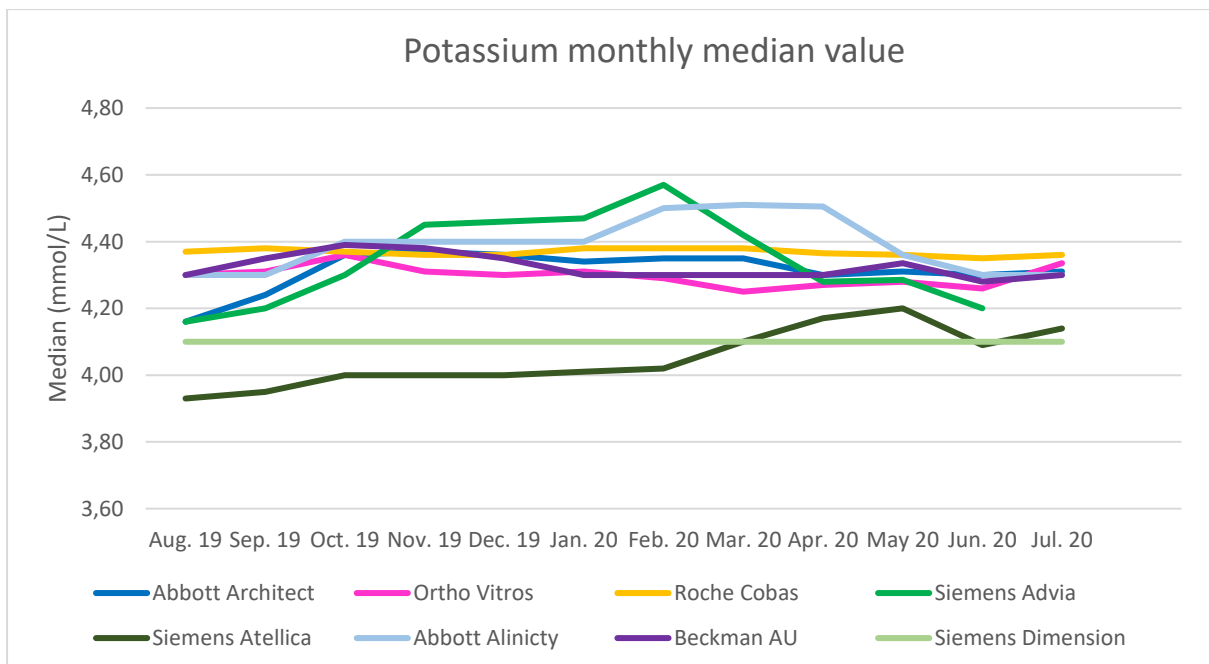
Phosphate



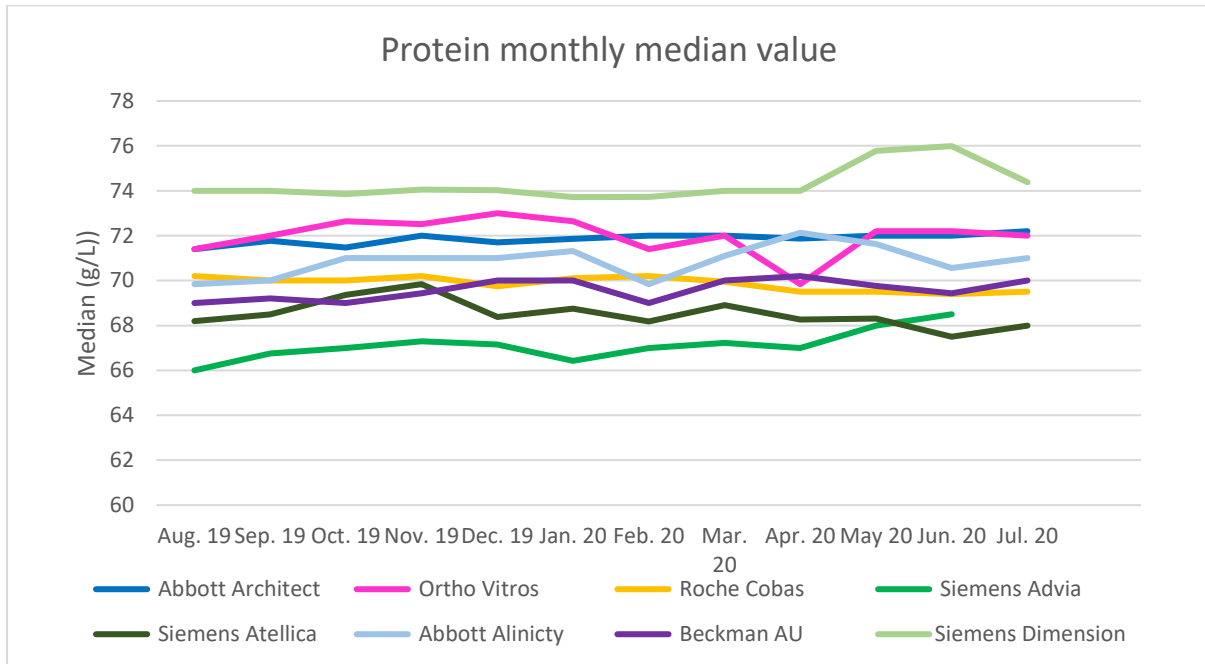
PLT



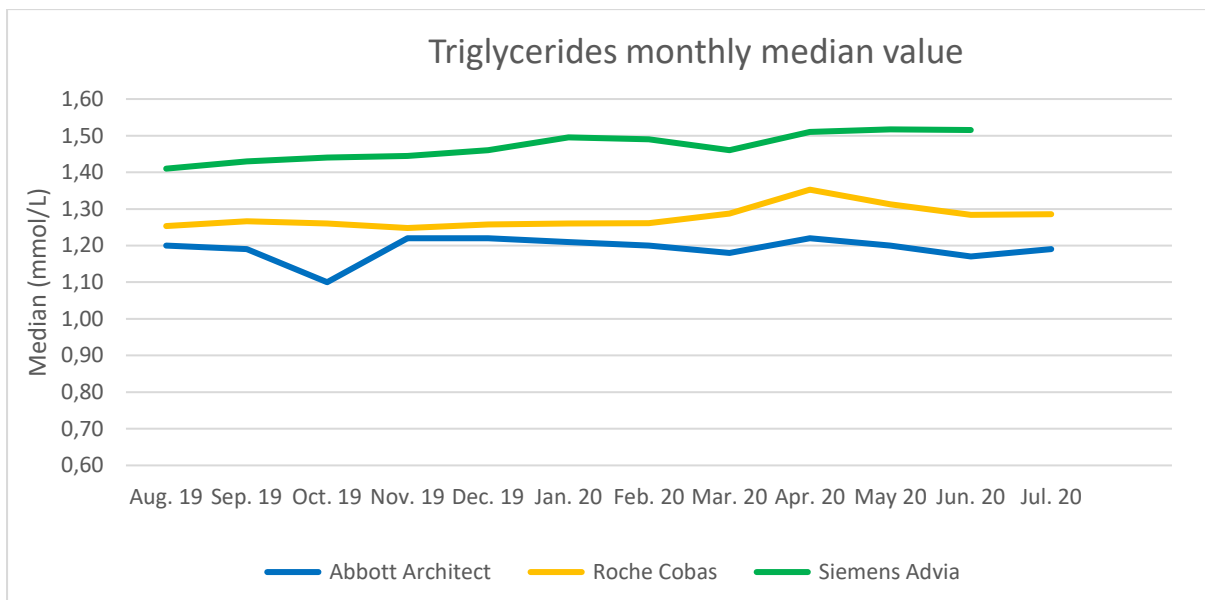
Potassium



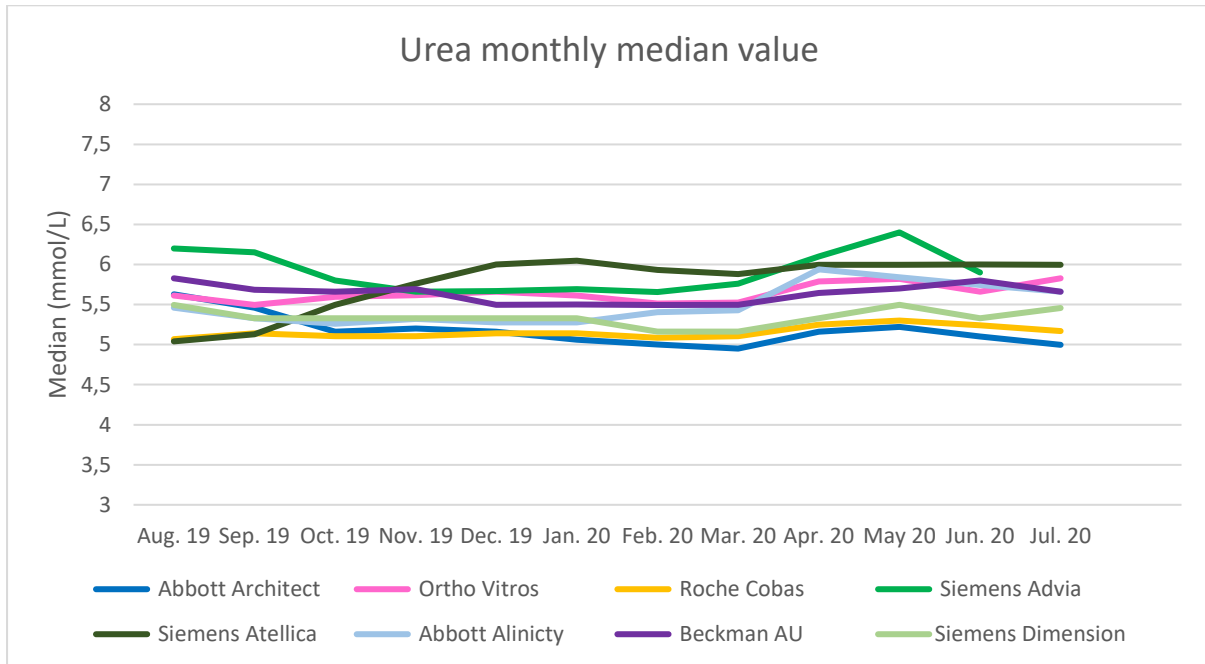
Protein (total)



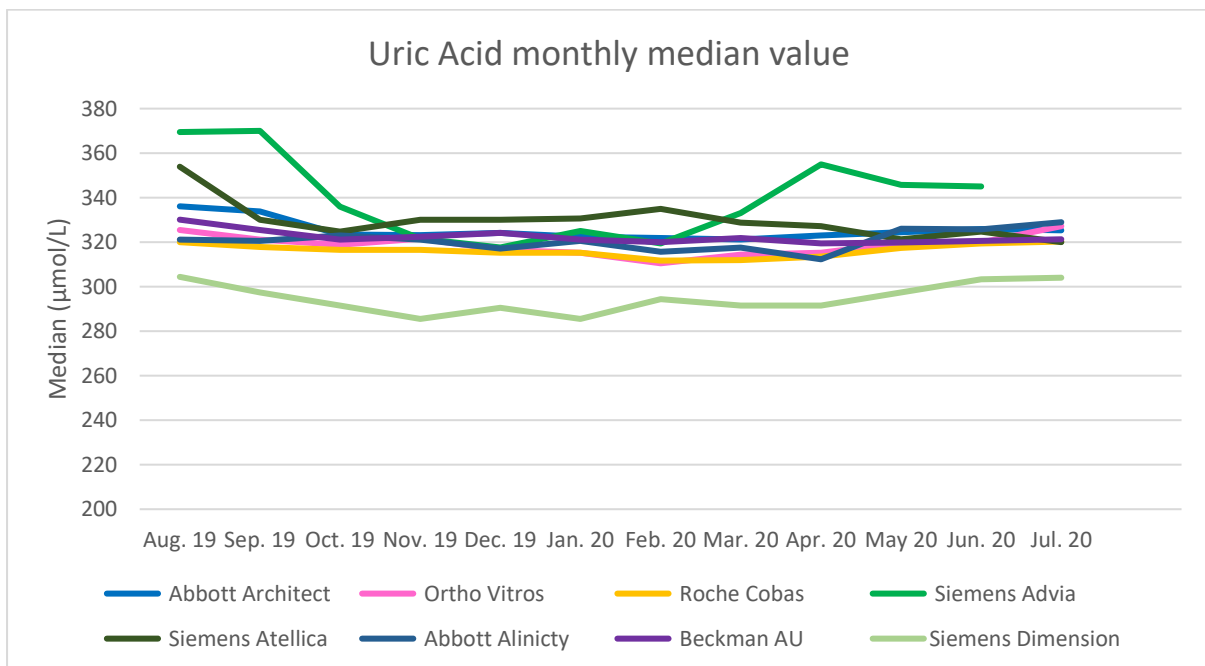
Triglycerides



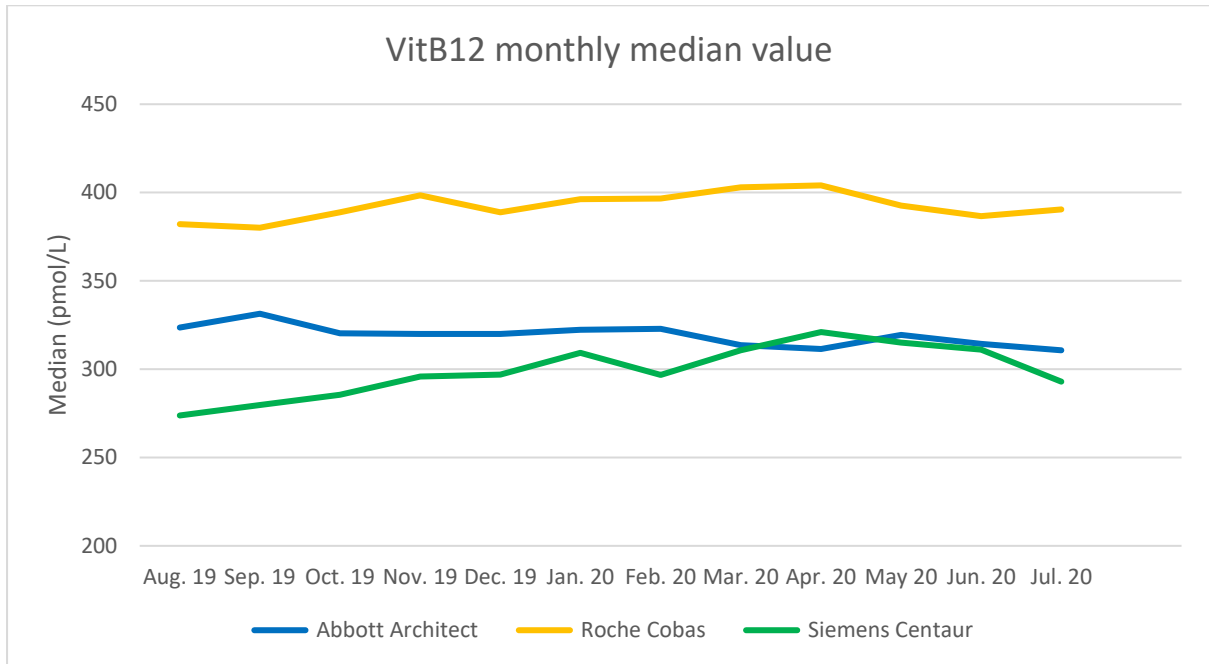
Urea



Uric acid



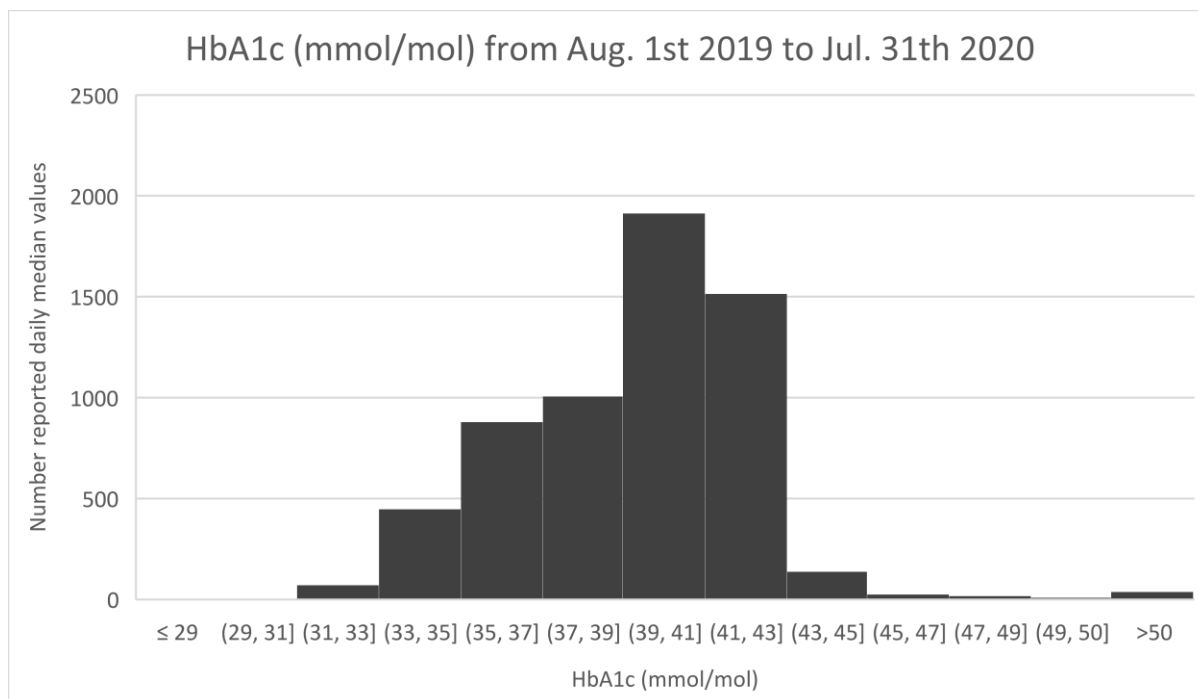
Vitamin B12



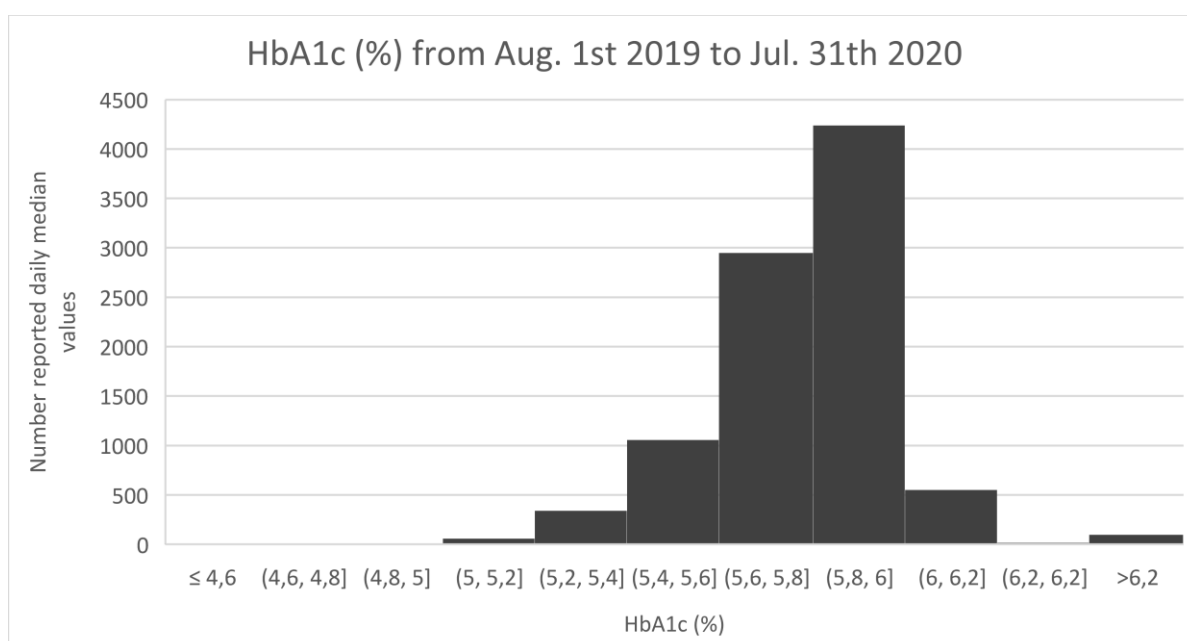
A short overview of the remaining analytes

For analytes not shown in figures, we have added histograms that show the distribution of all reported medians on the x-axis and the number of medians on the y-axis. These are mainly analytes with few participants today.

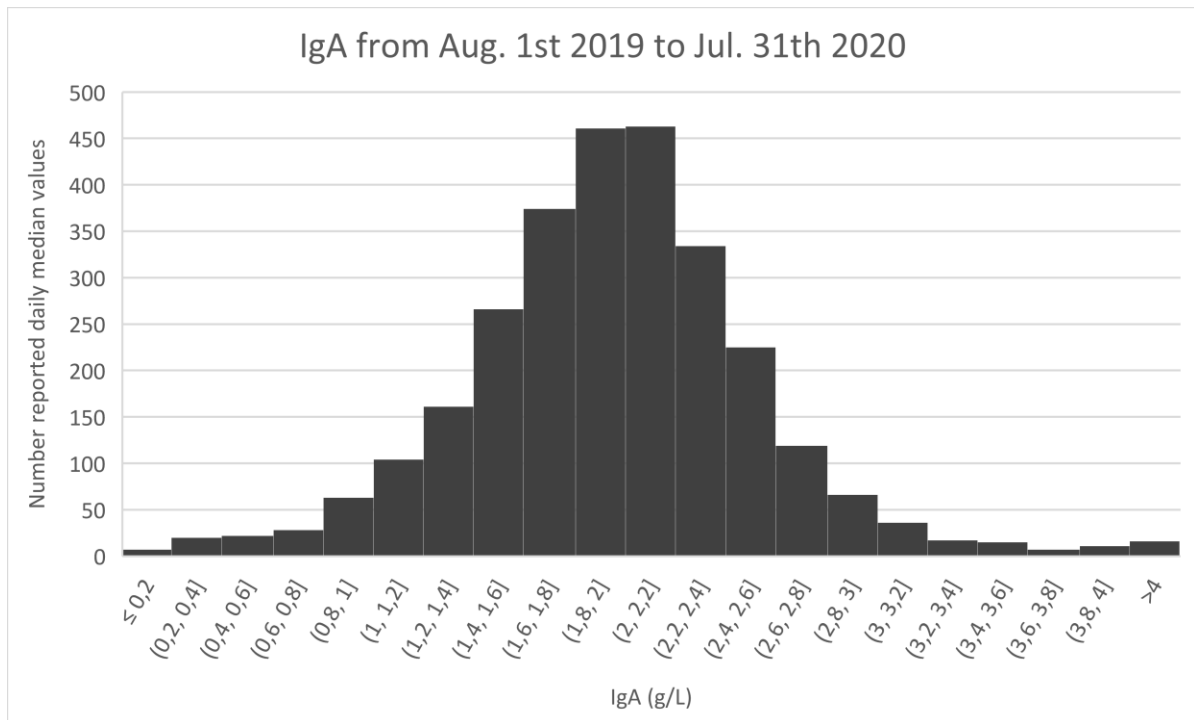
HbA1c (mmol/mol)



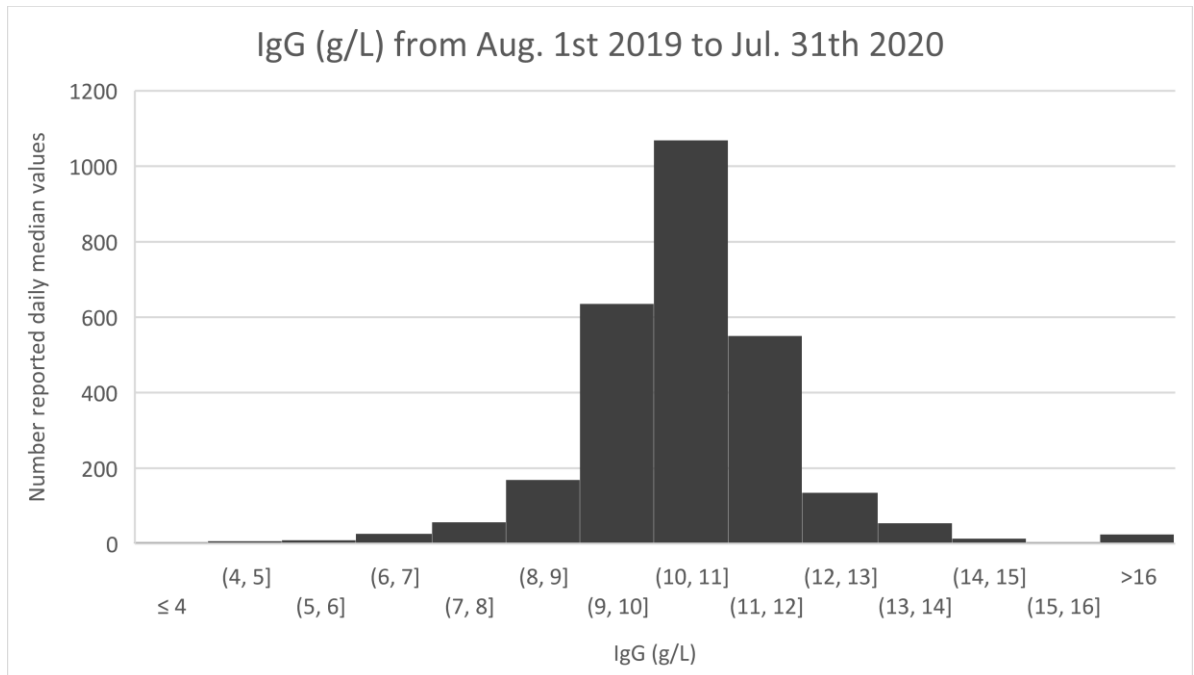
HbA1c (%)



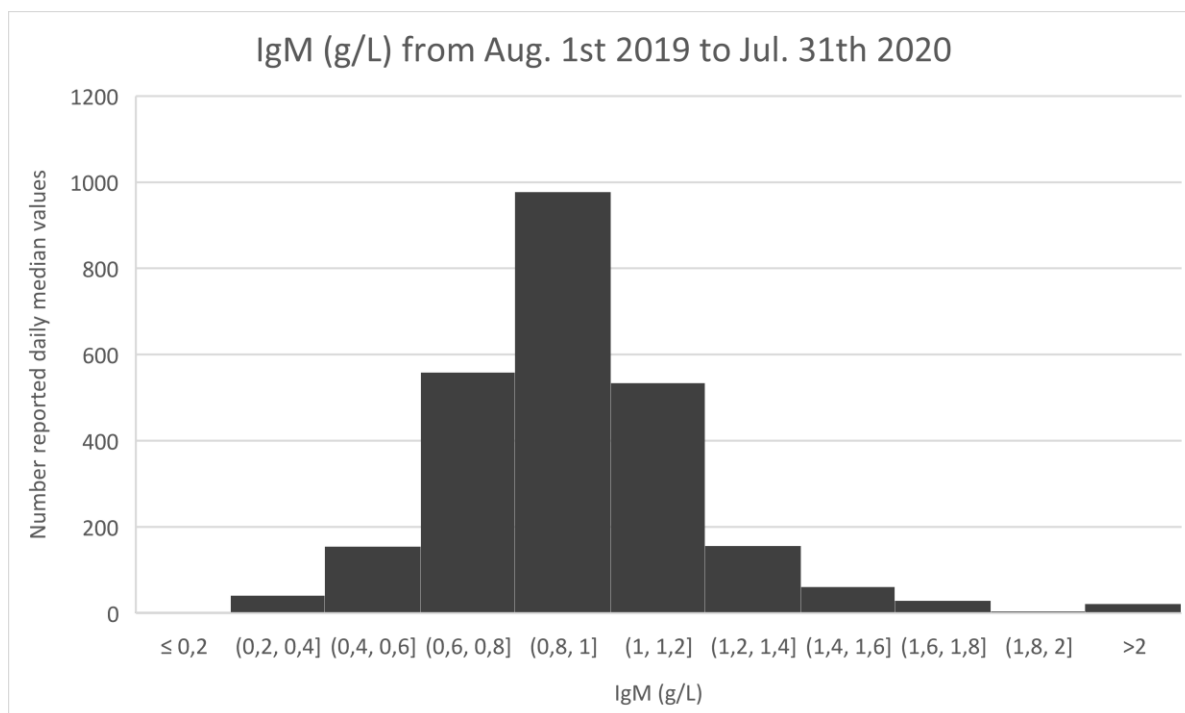
IgA



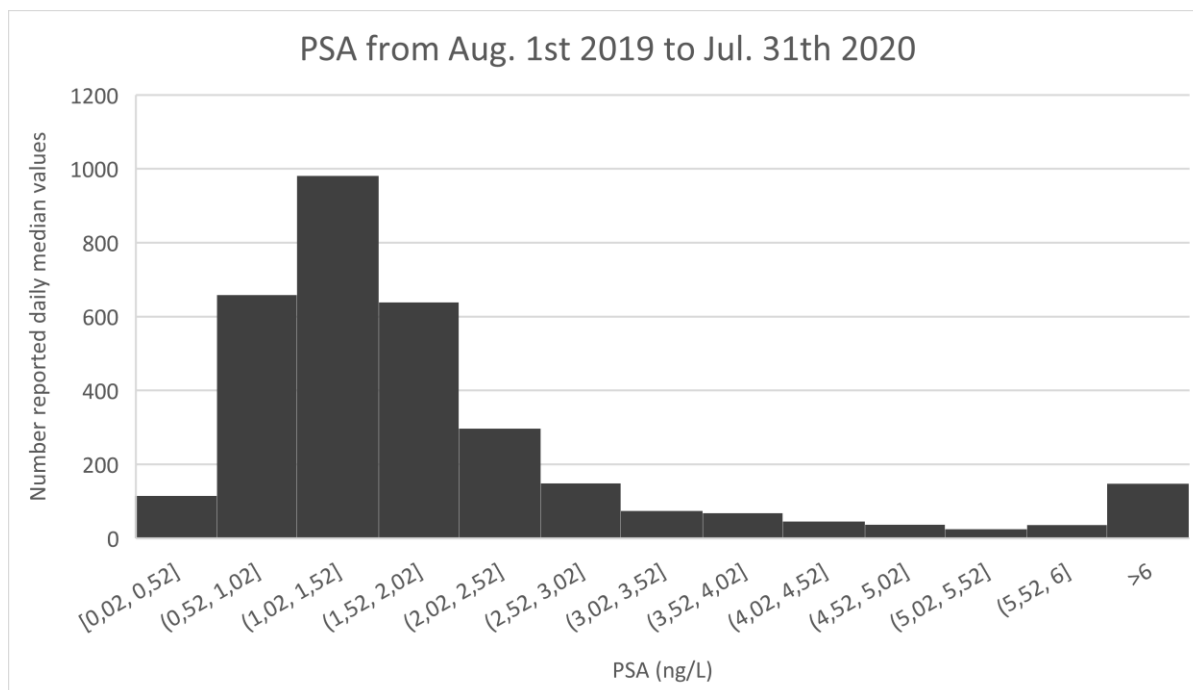
IgG



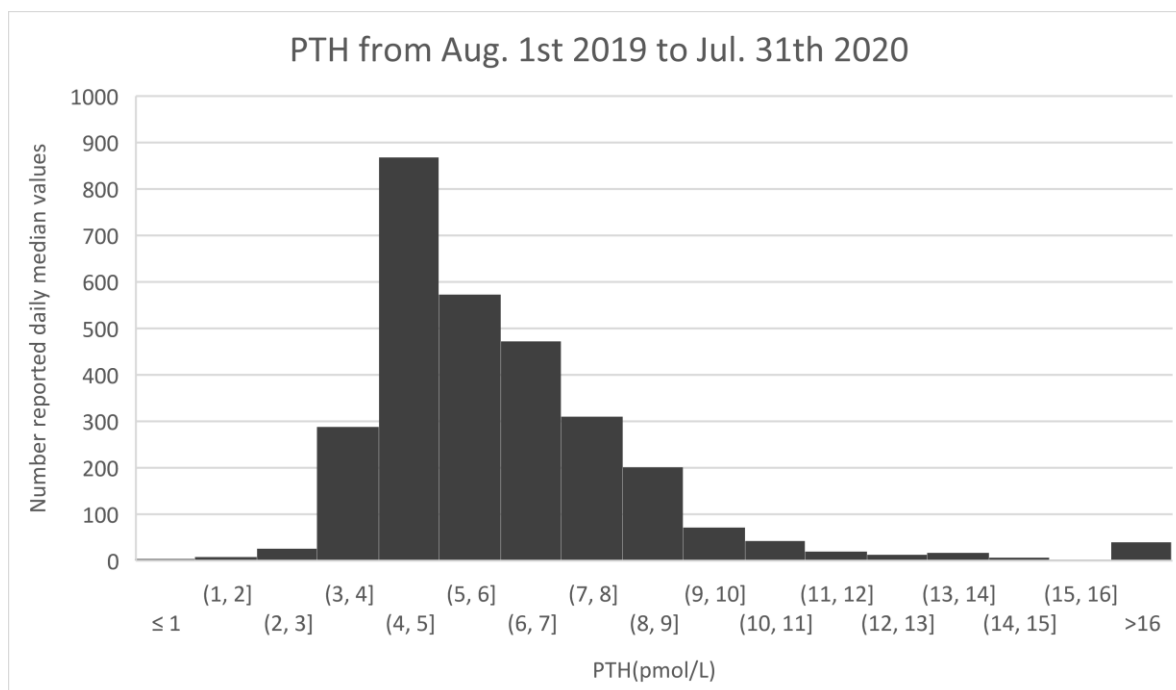
IgM



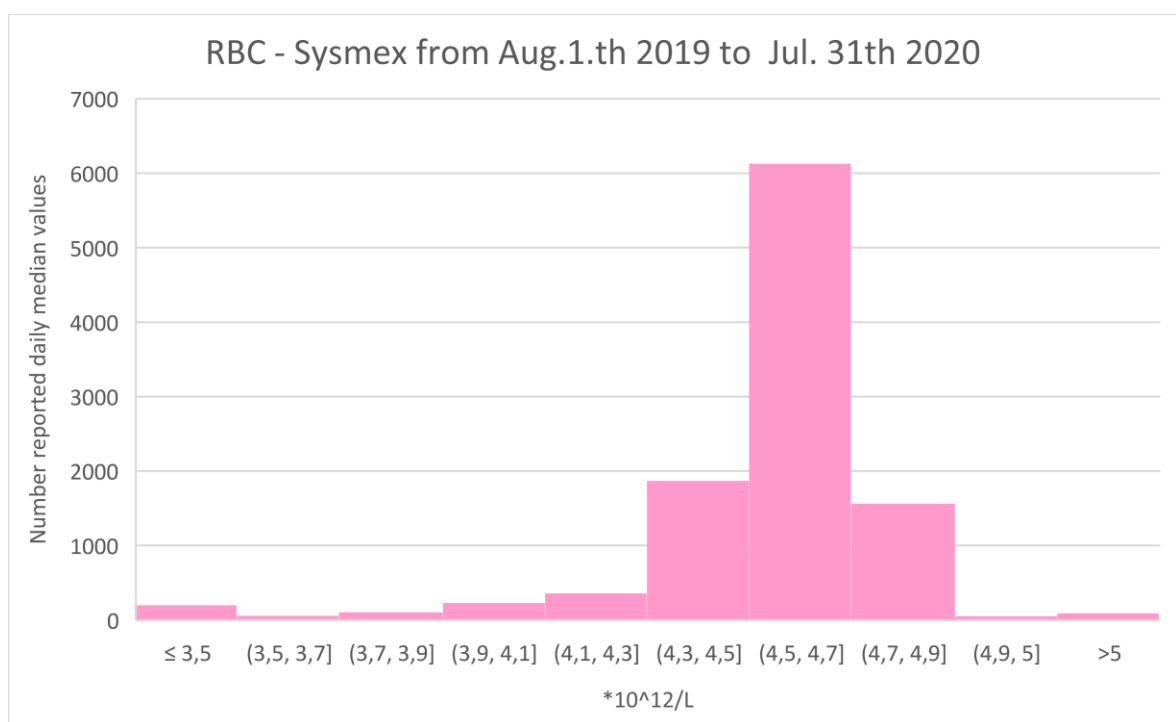
PSA

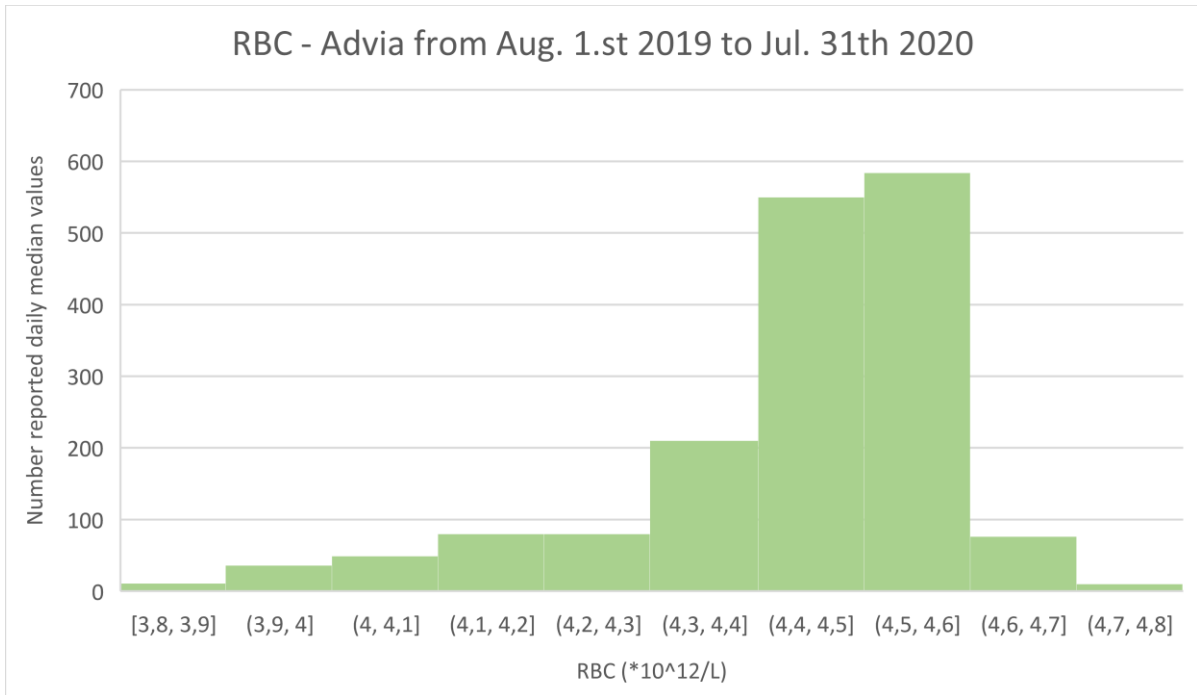


PTH

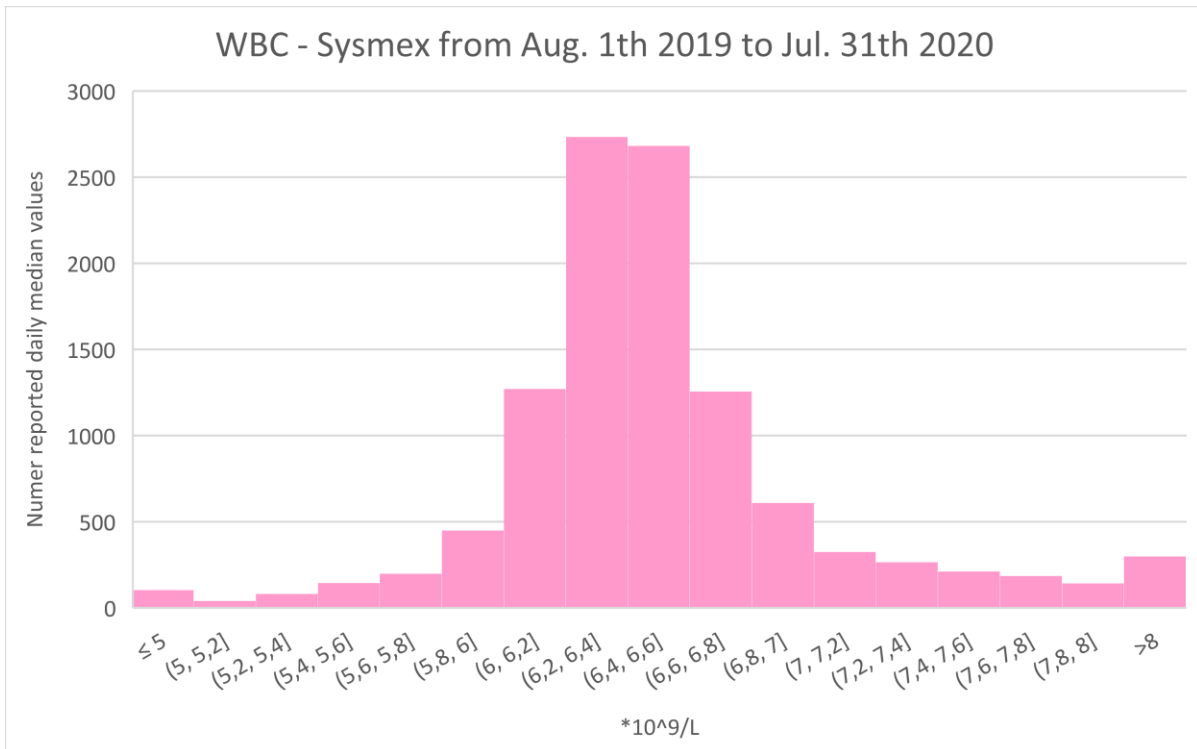


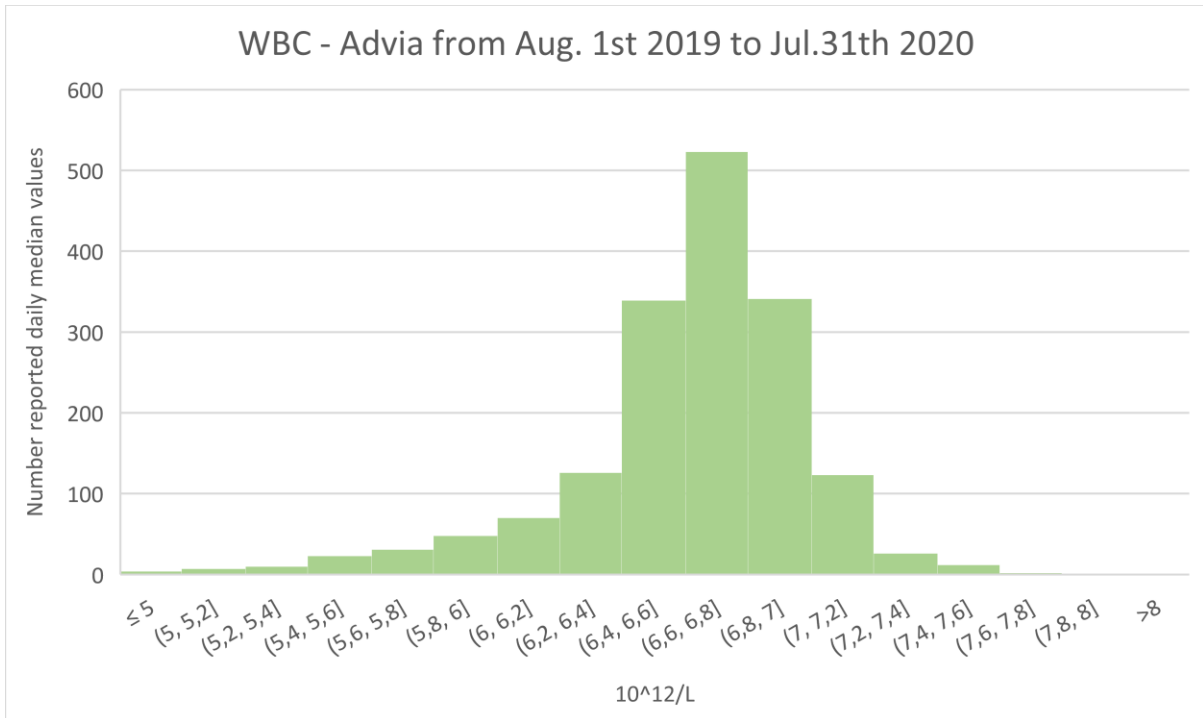
RBC





WBC





1. De Grande LA, Goossens K, Van Uytfanghe K, Stockl D, Thienpont LM. The Empower project - a new way of assessing and monitoring test comparability and stability. *Clin Chem Lab Med*. 2015;53(8):1197-204.
2. Goossens K, Van Uytfanghe K, Twomey PJ, Thienpont LM, Participating L. Monitoring laboratory data across manufacturers and laboratories--A prerequisite to make "Big Data" work. *Clin Chim Acta*. 2015;445:12-8.
3. Stepman HC, Stockl D, Twomey PJ, Thienpont LM. A fresh look at analytical performance specifications from biological variation. *Clin Chim Acta*. 2013;421:191-2.
4. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Martensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. *Scand J Clin Lab Invest*. 2004;64(4):271-84.
5. Adeli K, Higgins V, Seccombe D, Collier CP, Balion CM, Cembrowski G, et al. National Survey of Adult and Pediatric Reference Intervals in Clinical Laboratories across Canada: A Report of the CSCC Working Group on Reference Interval Harmonization. *Clin Biochem*. 2017;50(16-17):925-35.
6. Schumann G, Klauke R, Canalias F, Bossert-Reuther S, Franck PF, Gella FJ, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. Part 9: reference procedure for the measurement of catalytic concentration of alkaline phosphatase International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Scientific Division, Committee on Reference Systems of Enzymes (C-RSE) (1)). *Clin Chem Lab Med*. 2011;49(9):1439-46.
7. Braga F, Frusciante E, Infusino I, Aloisio E, Guerra E, Ceriotti F, et al. Evaluation of the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories. *Clin Chem Lab Med*. 2017;55(3):e47-e50.
8. Jorn Erlandsen E, Randers E. Challenges in the measurement of plasma creatinine on the Roche cobas c702. *Scand J Clin Lab Invest*. 2018;78(6):490-5.
9. van der Hagen EAE, Weykamp C, Sandberg S, Stavelin AV, MacKenzie F, Miller WG. Feasibility for aggregation of commutable external quality assessment results to evaluate metrological traceability and agreement among results. *Clin Chem Lab Med*. 2020.
10. Lee SC, Lim LM, Chang EE, Chiu YW, Hwang SJ, Chen HC. Effect of differences in serum creatinine estimation methodologies on estimated glomerular filtration rate. *Singapore Med J*. 2019;60(9):468-73.
11. Refardt J, Sailer CO, Chifu I, Winzeler B, Schnyder I, Fassnacht M, et al. The challenges of sodium measurements: indirect versus direct ion-selective method. *Eur J Endocrinol*. 2019;181(2):193-9.
12. De Grande LAC, Van Uytfanghe K, Reynders D, Das B, Faix JD, MacKenzie F, et al. Standardization of Free Thyroxine Measurements Allows the Adoption of a More Uniform Reference Interval. *Clin Chem*. 2017;63(10):1642-52.
13. Thienpont LM, Van Uytfanghe K, De Grande LAC, Reynders D, Das B, Faix JD, et al. Harmonization of Serum Thyroid-Stimulating Hormone Measurements Paves the Way for the Adoption of a More Uniform Reference Interval. *Clin Chem*. 2017;63(7):1248-60.