General Flagger Report

CONTENT

Introduction

Project status

Flagger stability limits

Percentiler – Flagger link

Conclusion

Introduction

The laboratory medicine community is concerned about analytical performance, however, struggles since long with defining specifications to assess quality. First laboratory inter-comparisons were validated by use of arbitrary specifications inferred from expert advice. Later on, several scientific approaches were developed based on the biological variation of the analytes. These were followed by many other models, including specifications derived from questionnaires to clinicians and requirements for decisions in specific clinical situations. A milestone was the Stockholm conference, defining a 5-level hierarchy for establishing analytical performance specifications, with specifications derived from “clinical situations” at the top. Unfortunately, little progress was made with implementing generally accepted “numbers”. This gave rise to holding a conference that addressed the situation 15 years later. In principle, the original proposal was re-iterated, however, the 5-level hierarchy was streamlined into 3 levels still with the clinically derived specifications at the top. Unfortunately, the clinical model gives only very few generally accepted numbers for routine laboratory practice. Therefore, specifications based on biological variation are still the most widely used ones. Only, there is no consensus about the actually desired numbers (optimum, desirable, minimum) and those that really can be applied in practice. Because of that, we expanded the ideas of Klee and investigated the effect of analytical instabilities (in fractions of the biological variation) on surrogate medical decisions, for which we used the flagging of results that exceed cut-off- or reference interval limits. Based on this theoretical framework, we developed 2 new web-based applications, the so-called Percentiler and Flagger. They intend to enable laboratories to monitor the stability of their analytical performance (Percentiler) and flagging rate (Flagger) directly from their own testing results for patients. To this end, the Percentiler and Flagger databases assemble respectively instrument-specific daily medians calculated from outpatient results and hypo/hyper flagging rates. The latter are the daily number of results (expressed as percentage relative to the total number) that are automatically flagged by the Laboratory Information System or middleware, when they are either lower or higher than the locally used decision limits. Combining both applications has the chance to bridge the medium hierarchy level (biological variation) with the top one (clinical outcome) to derive quality specifications. This may help individual laboratories with defining realistic but ambitious analytical performance specifications that apply for their own local situation.

Here, we report on our experience from the combined use of the Percentiler and Flagger. We put special emphasis on whether the observations realistically demonstrate the effect of the analytical performance on the flagging rate and whether the specifications set for the Flagger are fit-to-purpose in combination with those used in the Percentiler.

EMPOWERIVD●GLOBE

General Flagger Report 02/05/2017

Project status (May 2017)

Number of instruments for the clinical chemistry analytes

	Instruments per peer group	
Peer Group	**Percentiler**	**Flagger**
Abbott Architect	23	7
Beckman AU	12	11
Beckman DxC	6	2
Ortho Vitros	22	3
Roche Cobas	162	89
Roche Integra	3	3
Roche Modular	8	6
Siemens Advia	16	4
Siemens Dimension	10	3
Total	262	128

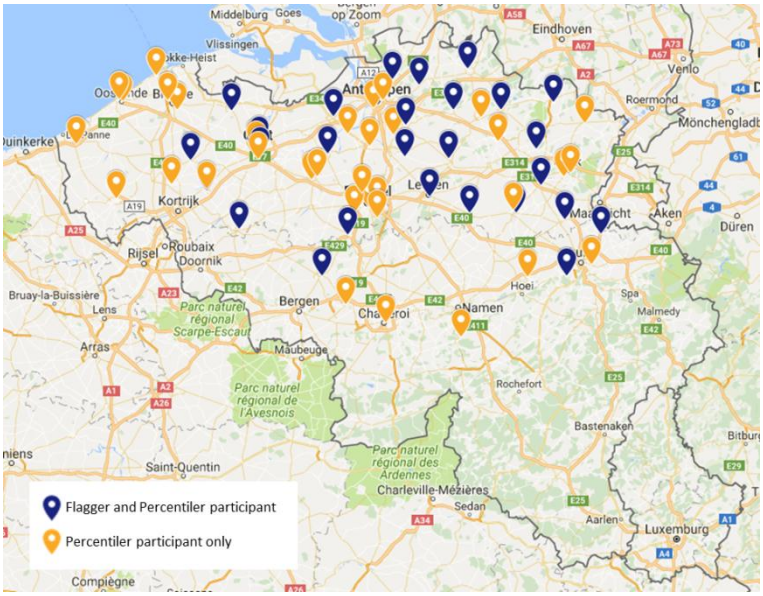
Number of instruments for FT4 and TSH

	Instruments per peer group	
Peer Group	**Percentiler**	**Flagger**
Abbott Architect	23	7
Beckman DxI	11	10
Roche Cobas ElecSys	78	32
Ortho Vitros	11	2
Siemens Advia Centaur	25	14
Siemens Dimension	7	3
Total	155	38

Note these numbers are estimates and vary for the different analytes.

Geographical distribution

Geographical distribution of the Percentiler and Flagger participants in Belgium:



Geographical distribution of the Percentiler and Flagger participants worldwide:

Flagger stability limits

The table below shows the specifications used in the Flagger and Percentiler application for the 22 examined analytes. The table reiterates the relation of the specifications for analytical performance to the desirable bias from the biological variation. The specifications for both applications are limits that should not be violated, but the percentage specifications for the Flagger are applied relatively to the long-term flagging rate observed for the individual laboratory. For example, if a laboratory has for AST a long-term flagging rate of 10%, the limits are set at ± 3% (= 30% of 10%); however, in case the long-term flagging rate is lower that 3.3%, the limit is set to a minimum value of 1%. Note that upon setting the specifications for the Flagger, we aimed at 30%, however, there were reasons to use more stringent or looser specifications (range: 20% to 70%). We defined the most stringent specification (20%) for total-cholesterol and glucose, because these analytes are of paramount importance in public health policies (coronary artery disease and diabetes); we applied 30% for 12 analytes which was in agreement with the current state-of-the-art performance, while for the remaining 8 analytes we used higher specifications (50% and 70%), in particular, for the analytes with a quite small biological variation (calcium, magnesium, sodium).

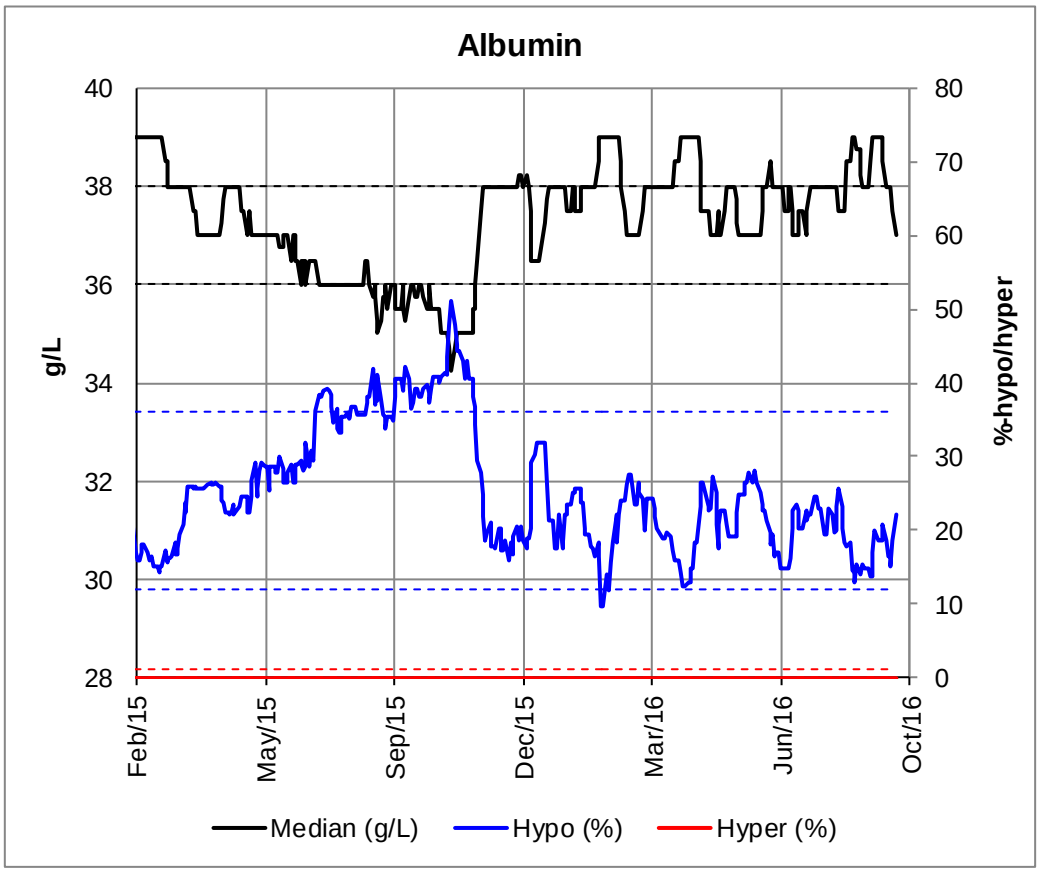
	Bias Limit from Biological variation (Absolute %)	Percentiler Limit (Absolute %)	Flagger Limit (Relative %)
ALB	1.43	2.4	50
ALKFOS	6.72	6.8	30
ALT	11.48	9.5	30
AST	6.54	6.5	30
BILTOT	8.95	10	30
CA	0.82	1.7	70
CHOL	4.1	3.8	20
CL	0.5	1	50
CRP	21.8	9.6	30
GGT	11.06	9.1	30
GLUC	2.34	3.1	20
K	1.81	2.4	30
CREAT	3.96	3.9	30
LDH	4.3	4.6	30
MG	1.8	3	70
NA	0.23	0.7	70
P	3.38	4.4	50
PROT	1.36	1.4	50
UREA	5.57	6	30
URIC ACID	4.87	4.8	50
FT4	3.3	3.3	30
TSH	7.8	7.7	30

Below we demonstrate that restricting the increase in flagging rates to 20% is realistic for some very important analytes (total-cholesterol and glucose) despite the fact that their relatively low biological variation requires ambitious analytical stability limits (2.3% for glucose, for example). This contrasts with the phosphate case (analytical stability limit of 3.4%), where we needed to operate the Flagger with a 50% limit. We speculate that this is because manufacturers (and laboratories) give special attention to testing of analytes that are in the public or scientific focus which, unfortunately, is not the case for the phosphate test. We could apply 30% Flagger limits for most of the other tests which nicely corresponds to analytical stability limits derived from biological variation and, consequently, limits applied in the Percentiler. For analytes with low biological variation (albumin, chloride, total-protein, calcium, magnesium, sodium) we had to set the Flagger limits higher (50% or 70%). Note, we also had to apply a 50% limit for uric acid because of its seasonal variation (somewhat higher in the summer).

Percentiler – Flagger link

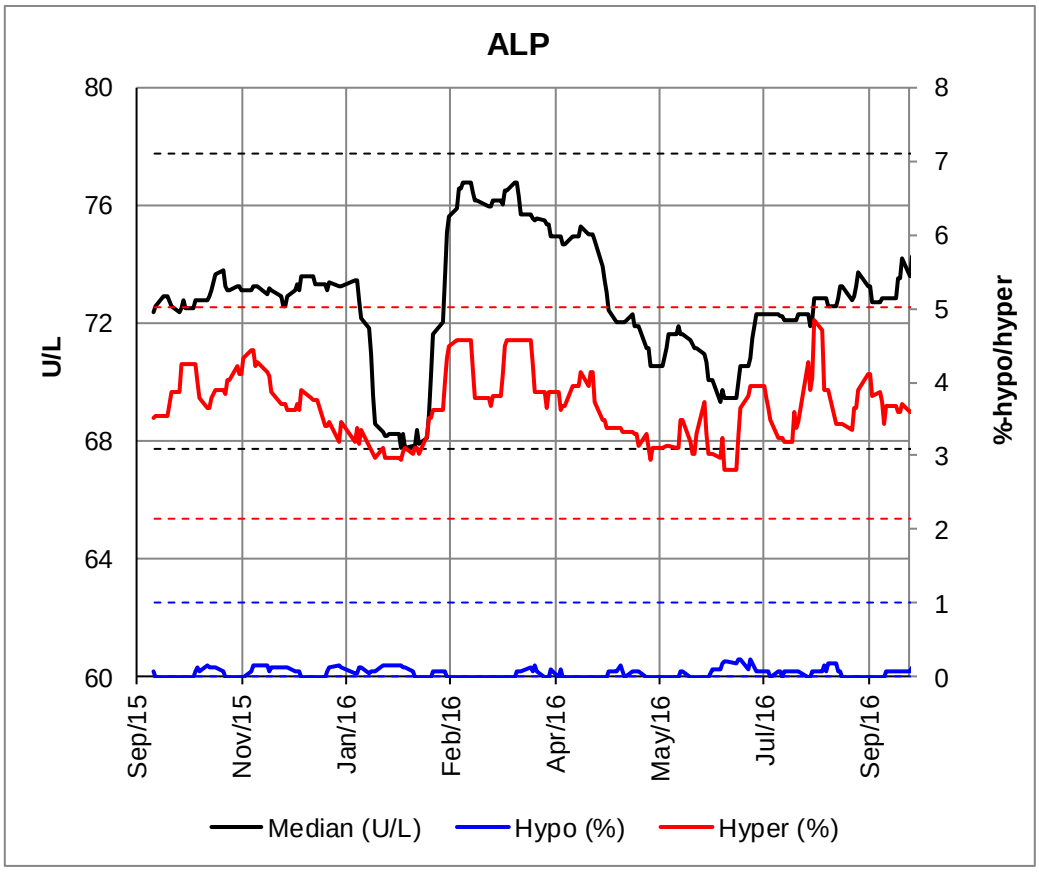
Our long-term experience with the Flagger and Percentiler applications allowed the investigation of the relationship between analytical performance specifications and surrogate medical decisions, i.e., flagging rates against locally used cut-offs or reference interval limits. However, it is to note that the Flagger and Percentiler applications monitor laboratory data at different concentration levels. Although this may cause that analytical variation observed in the Percentiler does not directly reflect changes in flagging rates, the majority of the selected examples show that the most prominent flagging rate (be it hypo or hyper) correlates quite well with analytical variations seen in the Percentiler application.

Albumin



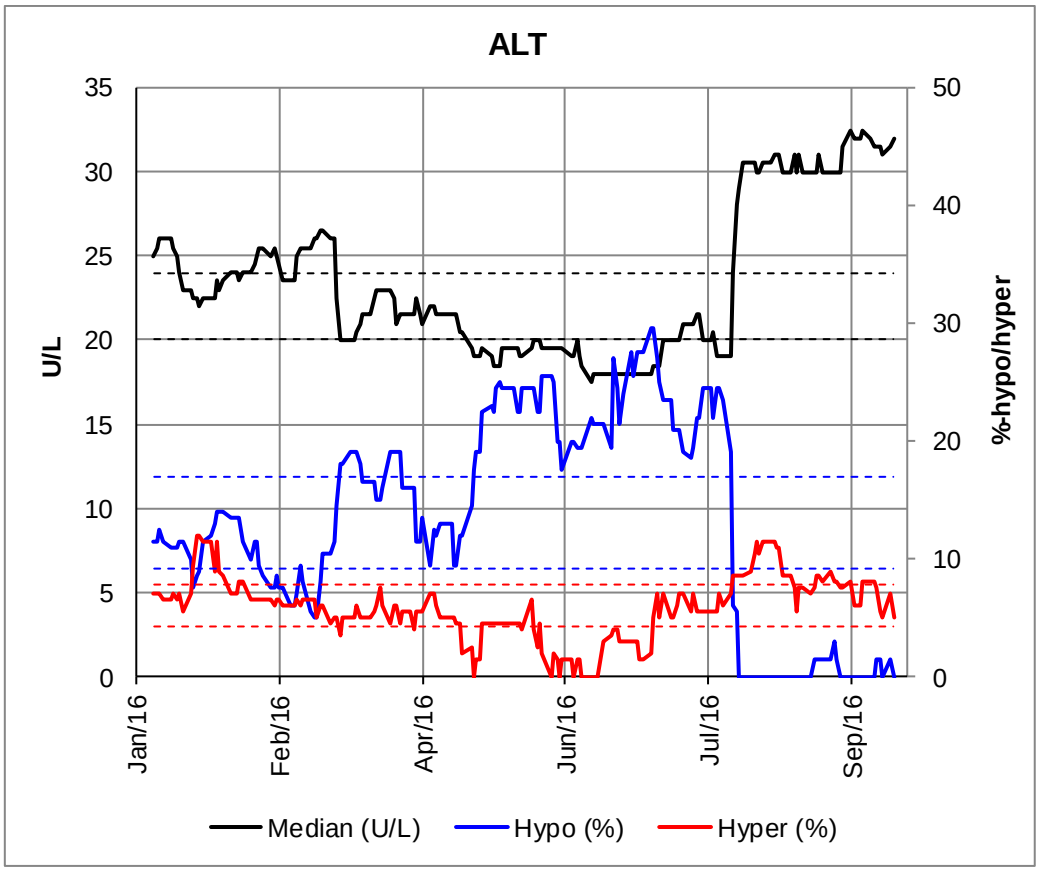
The figure shows a drift of albumin values from ~ 39 g/L to ~35 g/L, followed by a shift back to ~38 g/L. The hypo flagging rate increases from ~20% to ~42% and falls back to ~20%. This laboratory has, like most laboratories, no hyper flagging for albumin. Note, the Percentiler limit is 1 g/L (= 2.3%; “desirable” = 1.4%) and the Flagger limit is 50% of the long-term laboratory median.

ALP



ALP assays are among the most stable assays across laboratories and manufacturers. The figure here shows several moderate shifts of the alkaline phosphatase values, first the values drop from ~73U/L to ~68 U/L, second they go up to ~76 U/L, third they fall to ~70 U/L and fourth return to ~73U/L. The hypo flagging rate is very low, as mostly is the case. The hyper flagging rate varies concordantly with the analytical shifts between ~3 % to ~5%. Note, the Percentiler limit is 5 U/L (= 6.8%; “desirable” = 6.7%) and the Flagger limit is 30% of the long-term laboratory median.

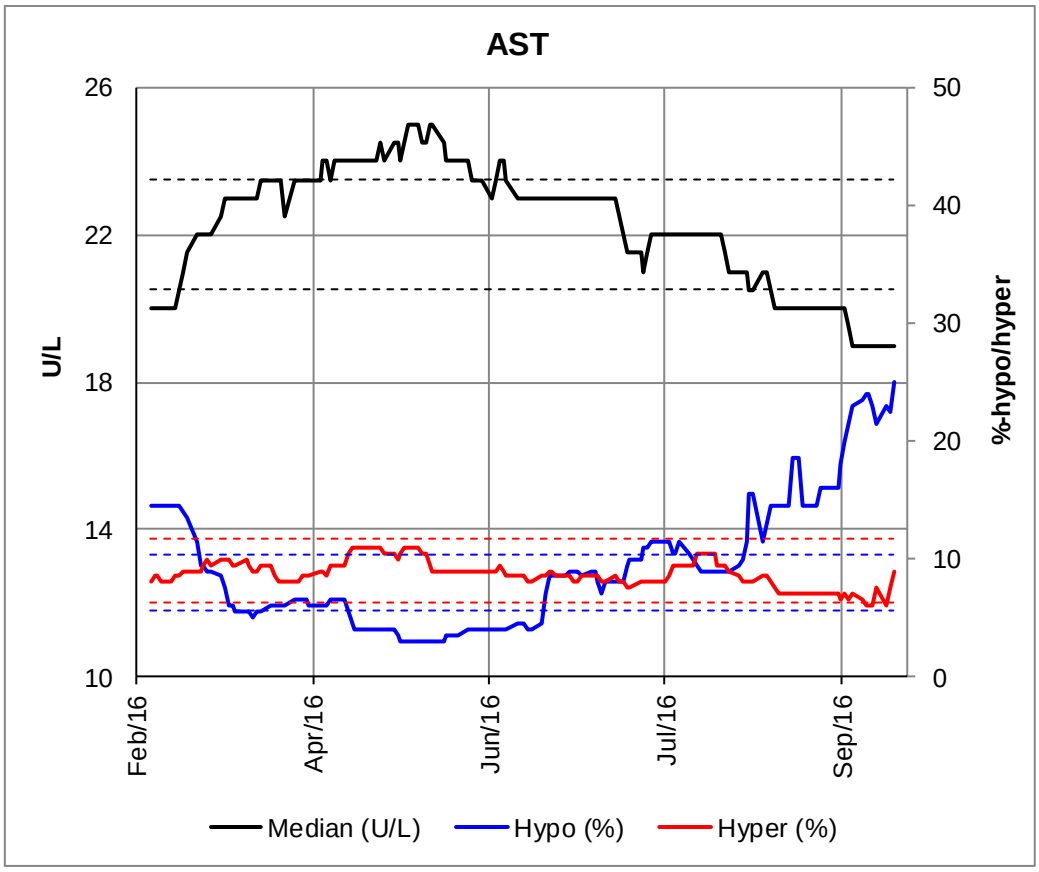
ALT



The figure shows a drift of the ALT values from ~24 U/L to ~18 U/L and a sharp shift to ~30 U/L. The hypo flagging rate increases from ~10% to ~28% and drops to 0%. The hyper flagging rate drops from ~7% to nearly 0% and increases, again, to ~8%. Note, the Percentiler limit is 2 U/L (= 9.5%; “desirable” = 11.5%) and the Flagger limit is 30% of the long-term laboratory median.

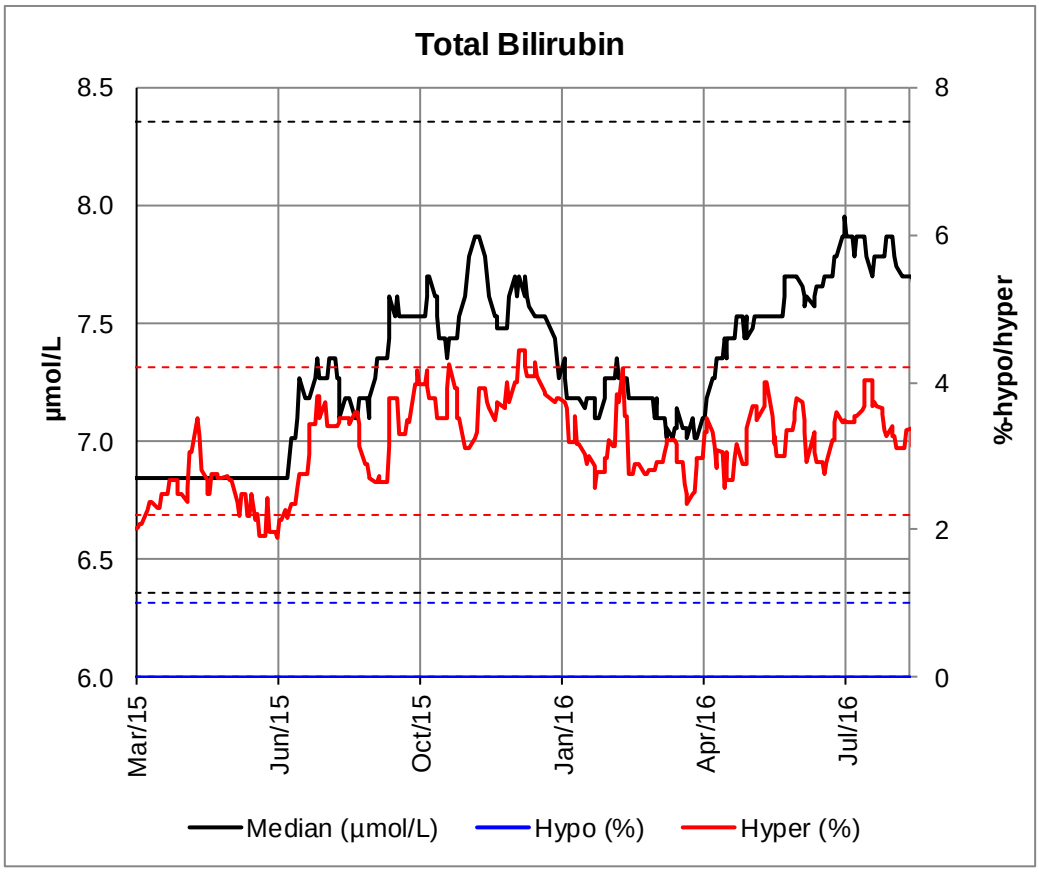
ALT tests, typically, were stable within the recommended limits, except the one of Ortho Clinical Diagnostics used to construct the figure and demonstrating that the analytical instability gave significant changes in flagging rates (hypo from 0% to 28%; hyper from 0% to 8%). Analytical stability in the order of 2 IU for concentrations in the reference interval could greatly support the utility of the ALT test for newer applications, such as early detection of metabolic changes (“metabolic syndrome”). In that connection, cut-offs for high-normal may be useful as most laboratories, only, flag very high results.

AST



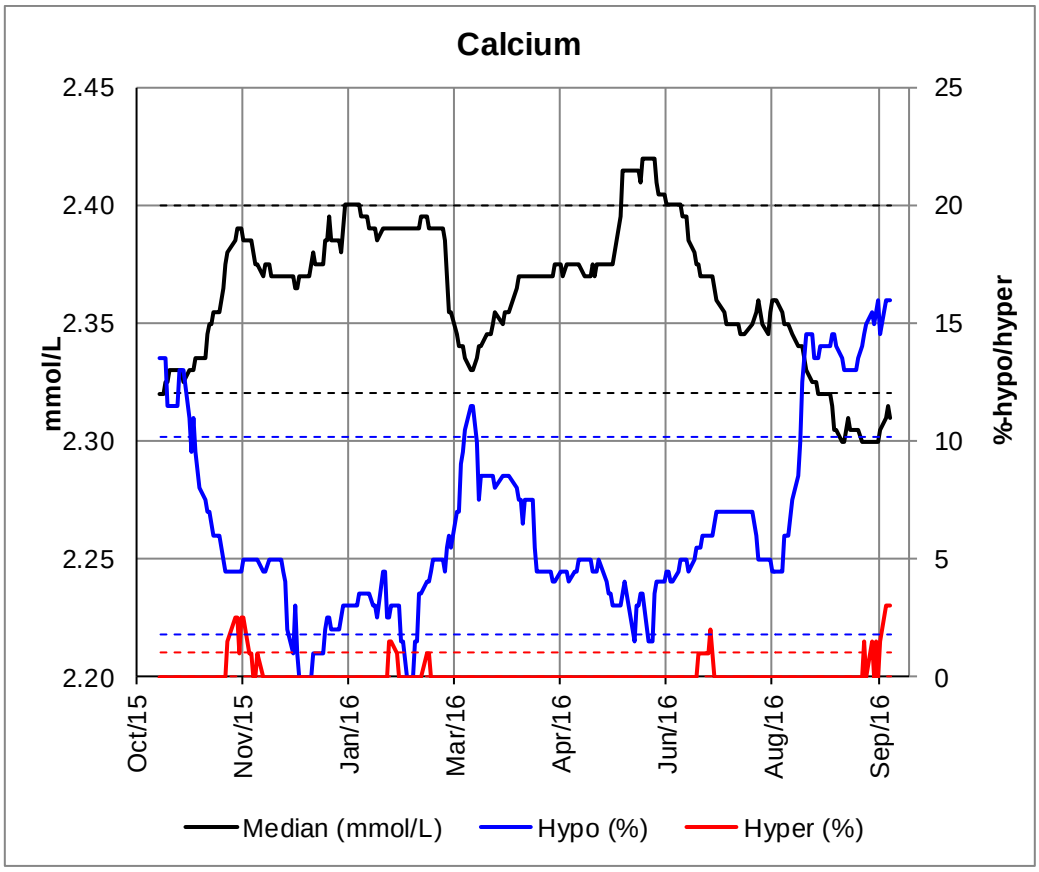
The figure shows AST values varying from ~20 U/L to ~24 U/L and back to ~20 U/L. The hypo flagging rate decreases concordantly with the analytical variation from ~13% to ~4% and increases again to ~22%. The hyper flagging rate slightly drops from ~8% to ~6%. Note, the Percentiler limit is 1.5 U/L (= 6.5%; “desirable” = 6.5%) and the Flagger limit is 30% of the long-term laboratory median.

Total-bilirubin



The total-bilirubin assay is among the most stable assays across manufacturers and laboratories; typically, the limits are never violated. The figure shows total bilirubin values shifting from ~6.8 µmol/L up to ~7.5 µmol/L; then they decrease to ~7.0 µmol/L followed by an increase to ~7.6 µmol/L. The hyper flagging rate is concordantly affected by both analytical shift as it goes from ~2.5% up to nearly 4% and comes back to ~3%. As expected, the hypo flagging rate is almost zero %. Note, the Percentiler limit is 1 µmol/L (= 10%; “desirable” = 9.0%) and the Flagger limit is 30% of the long-term laboratory median.

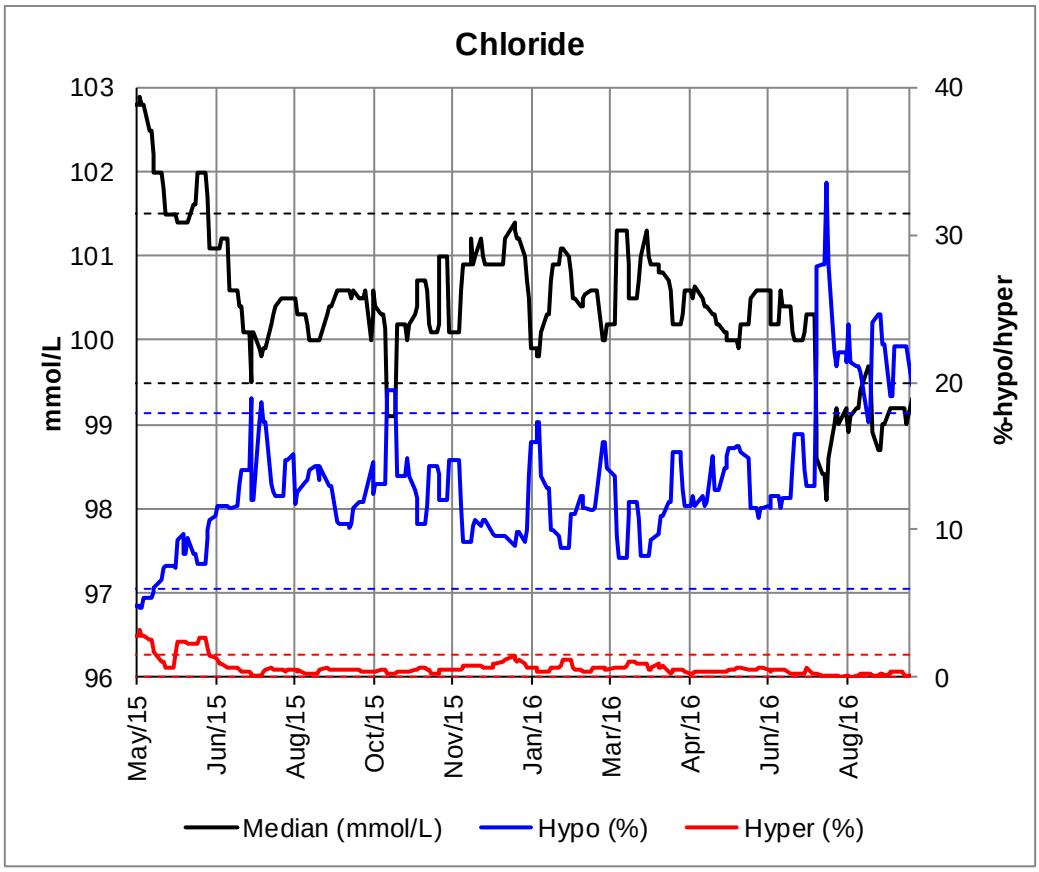
Calcium



The figure shows moderately varying calcium values, i.e. they first shift upwards from ~2.33 mmol/L to ~2.38 mmol/L then they fall back to 2.33 mmol/L, gradually increase a second time to ~2.42 mmol/L, to finally drop back to 2.30 mmol/L. The hypo flagging rate drops more strongly with the analytical shifts from ~13% to ~ 3%, then increases to ~9%, falls back to 4% to finally increase a second time to ~15%. The hyper flagging rate is very low so that no effects are observed. Note, the Percentiler limit is 0.04 mmol/L (= 1.7%; “desirable” = 0.8%) and the Flagger limit is 70% of the long-term laboratory median.

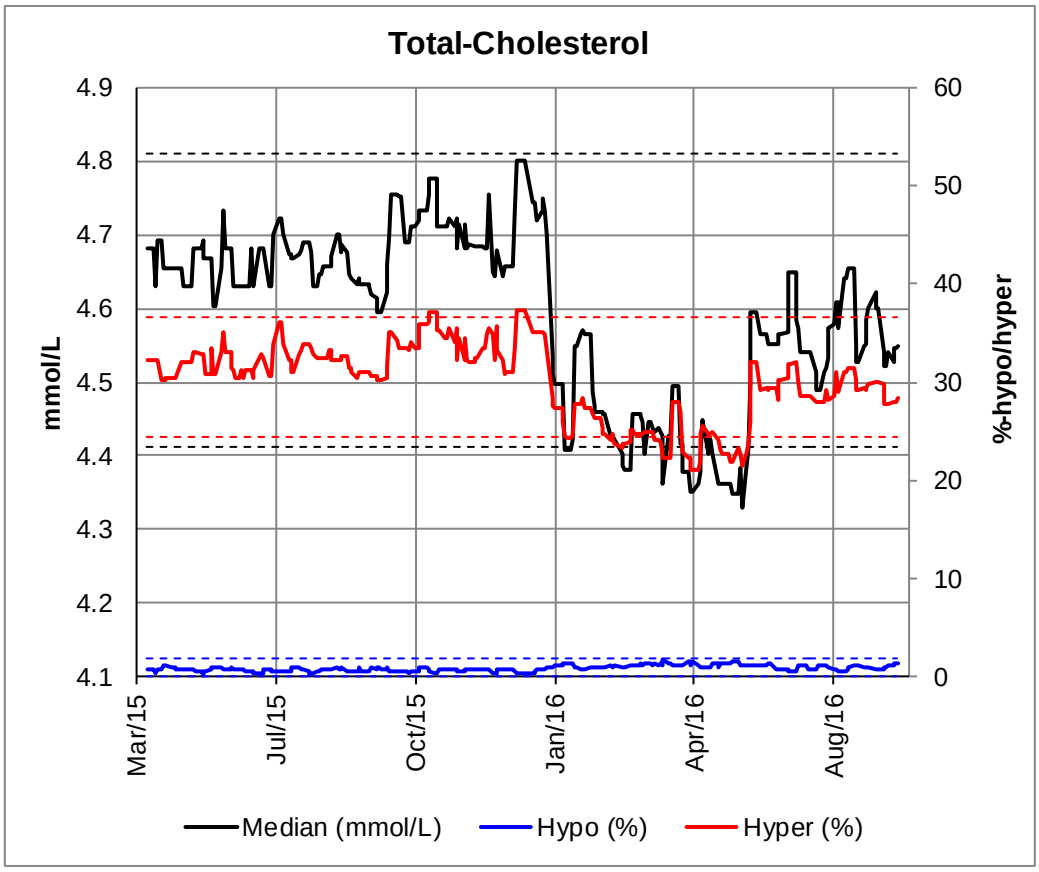
Calcium flagging rates are very much influenced by analytical instability because of the low biological variation. The hypo flagging rate was ~15% at 2.30 mmo/L, whereas it was only ~3% at 2.42 mmol/L in the example shown. This means that operating the calcium test with an instrument or laboratory bias of ~0.12 mmol/L could result in 5-fold different hypo flagging rates.

Chloride



The figure shows 2 shifts in the chloride values, i.e., from ~ 103 mmol/L to ~100.5 mmol/L, and to ~99 mmol/L. The hypo flagging rate is most influenced by both analytical shifts as it increases from ~5% to ~21%. The hyper flagging rate decreases by the first downward shift from ~2% to nearly 0%. Note, the Percentiler limit is 1 mmol/L (= 1.0%; “desirable” = 0.5%) and the Flagger limit is 50% of the long-term laboratory median.

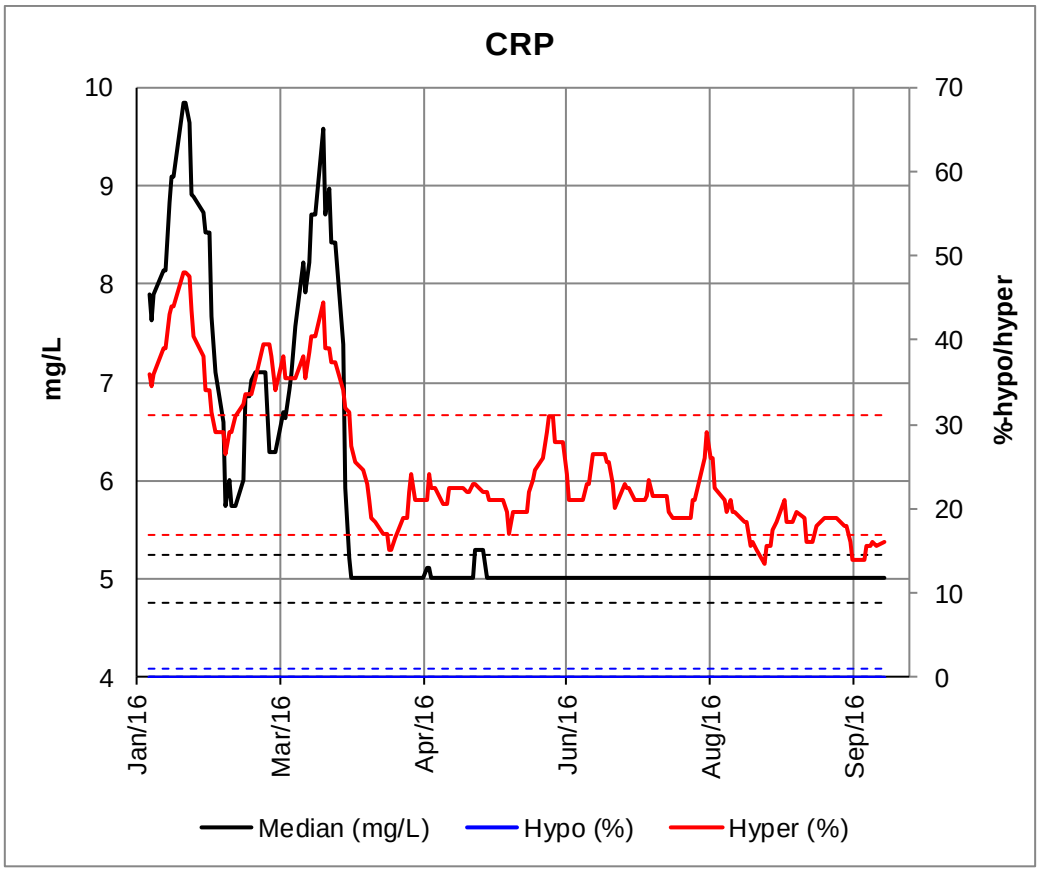
Total-Cholesterol



The figure shows a major shift in total cholesterol values from ~4.7 mmol/L to ~4.4 mmol/L. The hypo flagging rate is nearly unaffected and in the order of ~1%, while the hyper flagging rate decreases from ~36% to nearly 23%. Note, the Percentiler limit is 0.2 mmol/L (= 3.8%; “desirable” = 4.1%) and the Flagger limit is 20% of the long-term laboratory median.

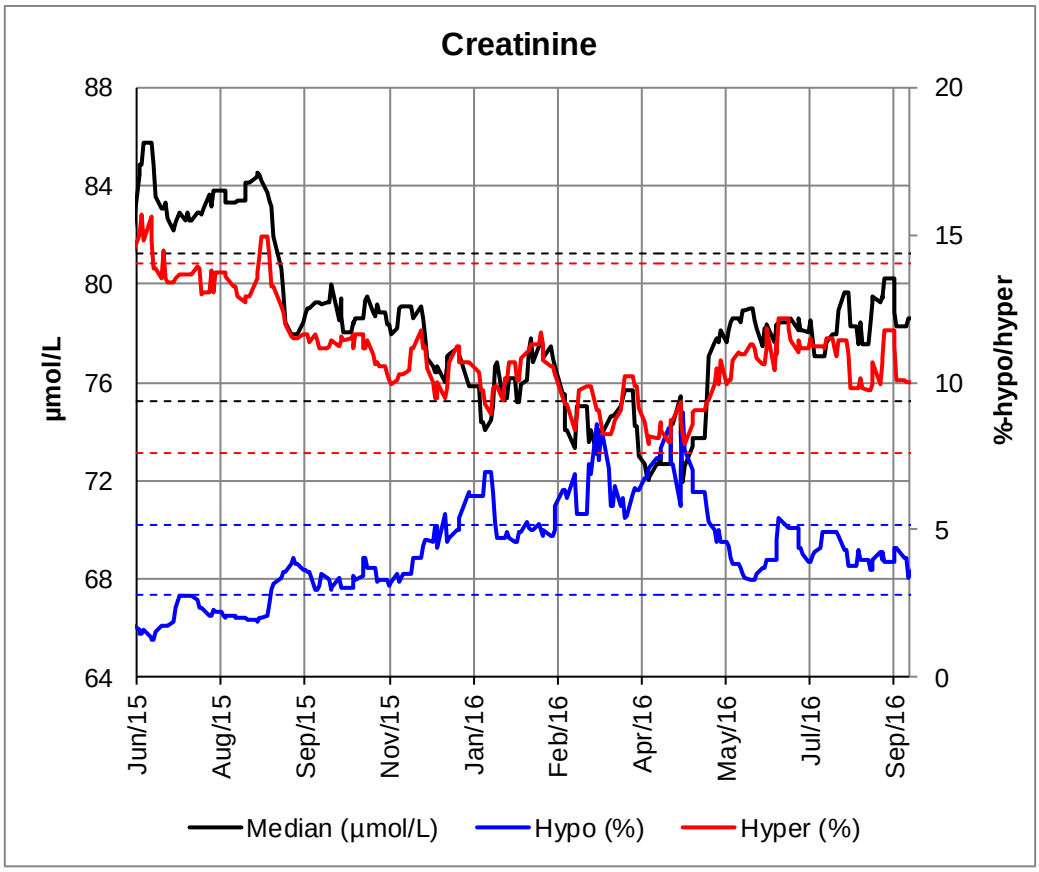
The total-cholesterol example demonstrates the maturity of the test. Certain laboratories were able to keep the test within the 20% Flagger limit for over 2 years, corresponding to an analytical stability within ±0.2 mmol/L.

CRP



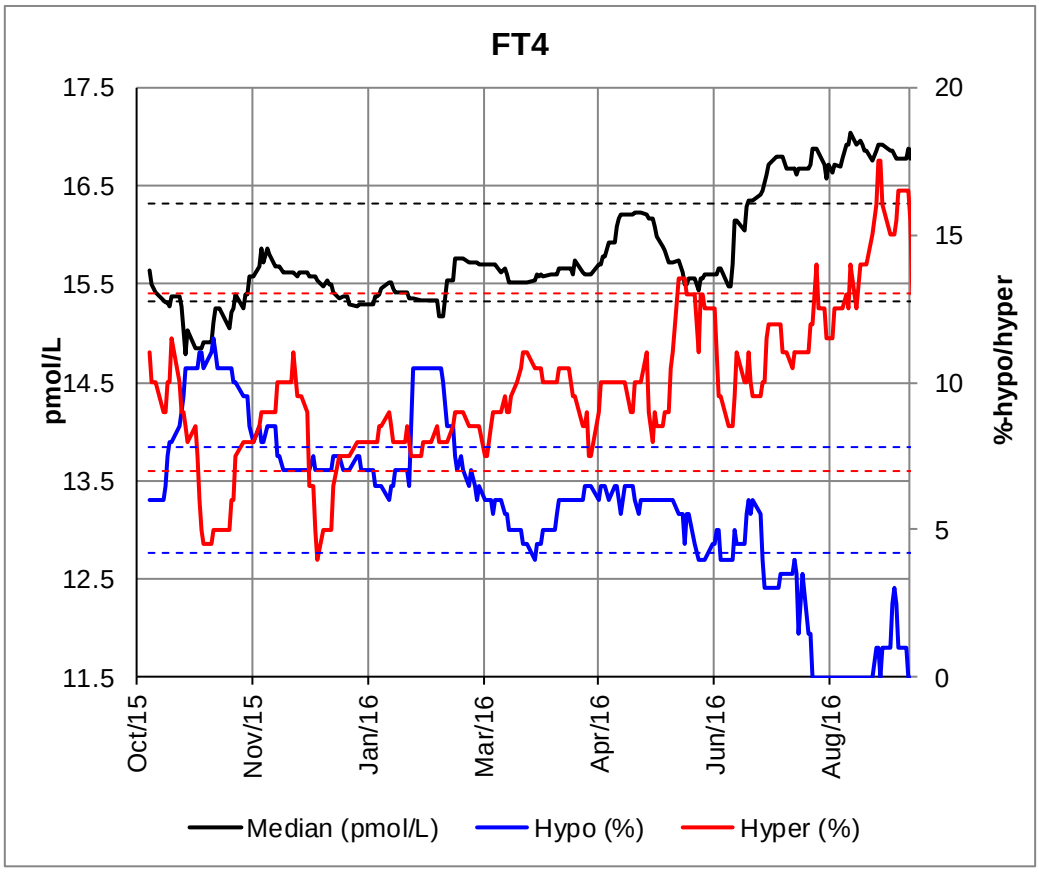
The figure shows 4 major shifts in the CRP values, i.e., from ~7.0 mg/L to ~ 9.9 mg/L, then down to ~5.8 mg/L, up to ~9.5 mg/Land down to ~5 mg/L. The hyper flagging rate is influenced by the analytical shifts as it first increases from ~35% to ~48%, then drops to ~28% to increase again to ~35-45%, and then decreases to ~15-20%. The hypo flagging rate is and remains zero as expected. Note, the Percentiler limit is 0.25 mg/L (= 9.6%; “desirable” = 21.8%) and the Flagger limit is 30% of the long-term laboratory median.

Creatinine



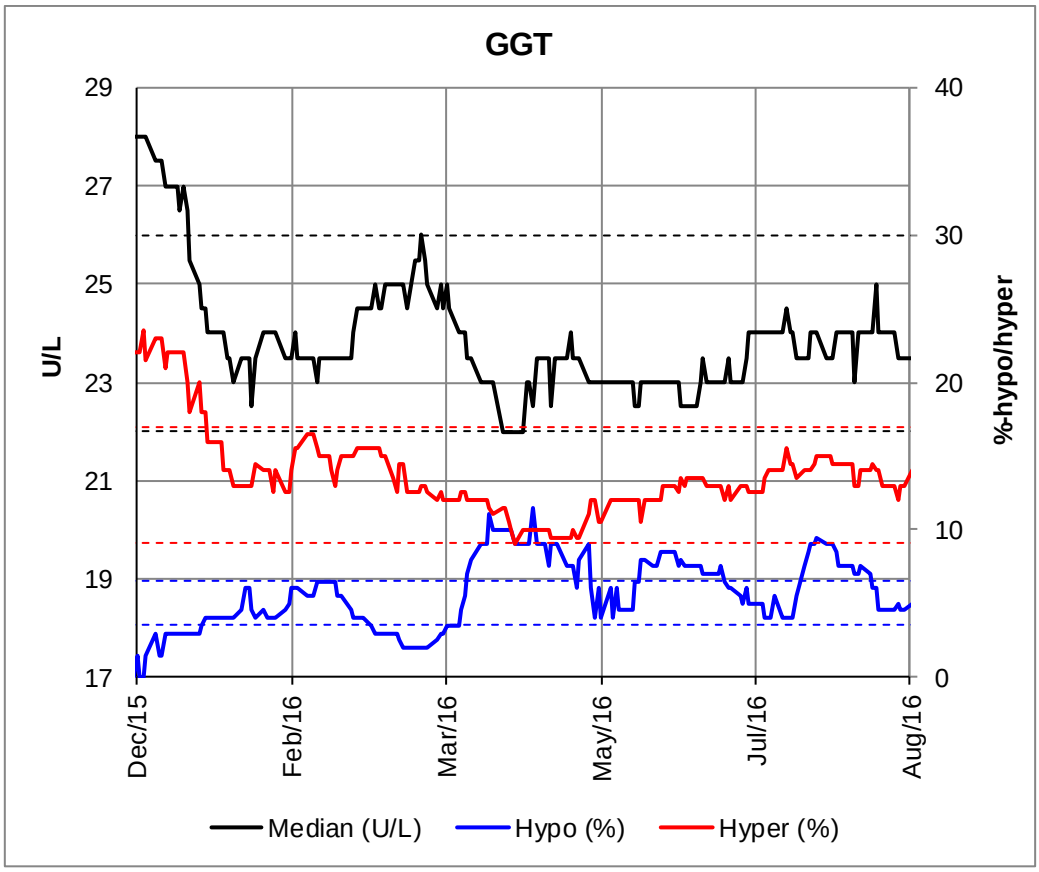
The figure shows drifting creatinine values from ~84 µmol/L down to ~74 µmol/L, followed by a shift to~78 µmol/L. The hypo flagging increases concordantly with the drift from ~3% to ~7.5% to normalize again to ~3%. The hyper flagging rate gradually decreases from ~15% down to nearly 8% and then normalizes to ~11%. Note, the Percentiler limit is 3 µmol/L (= 3.9%; “desirable” = 4.0%) and the Flagger limit is 30% of the long-term laboratory median.

FT4



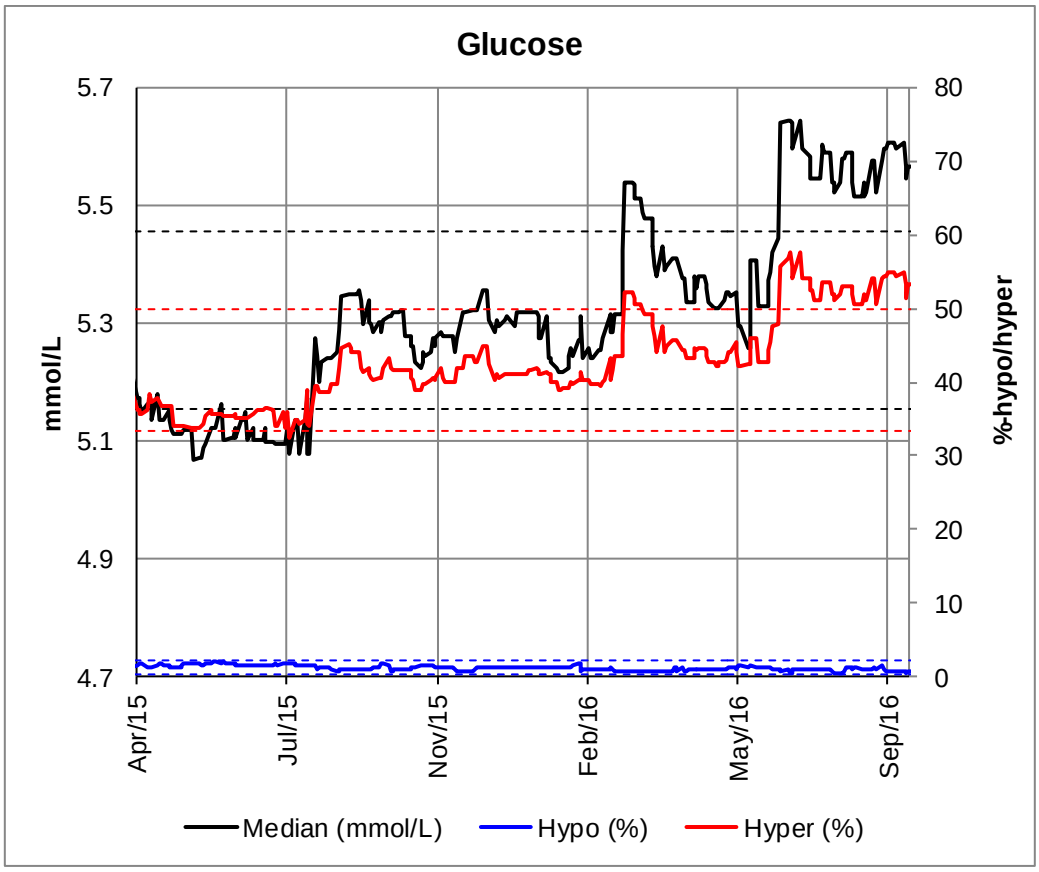
The figure shows FT4 values drifting and shifting from 15.5 pmol/L up to 17 pmol/L. The hypo flagging rates decrease from ~10% to ~0%, while the opposite happens for the hyper flagging rate (increases from ~8% to nearly 16%). Note, the Percentiler limit is 0.5 pmol/L (= 3.3%; “desirable” = 3.3%) and the Flagger limit is 30% of the long-term laboratory median.

GGT



The figure shows a downward shift in GGT values from ~28 U/L to ~24U/L. The hypo flagging rate increases concordantly with the analytical shift from ~ 3% to ~10%. The hyper flagging rate drops from ~22% to nearly 10%. Note, the Percentiler limit is 2 U/L (= 9.1%; “desirable” = 11.1%) and the Flagger limit is 30% of the long-term laboratory median.

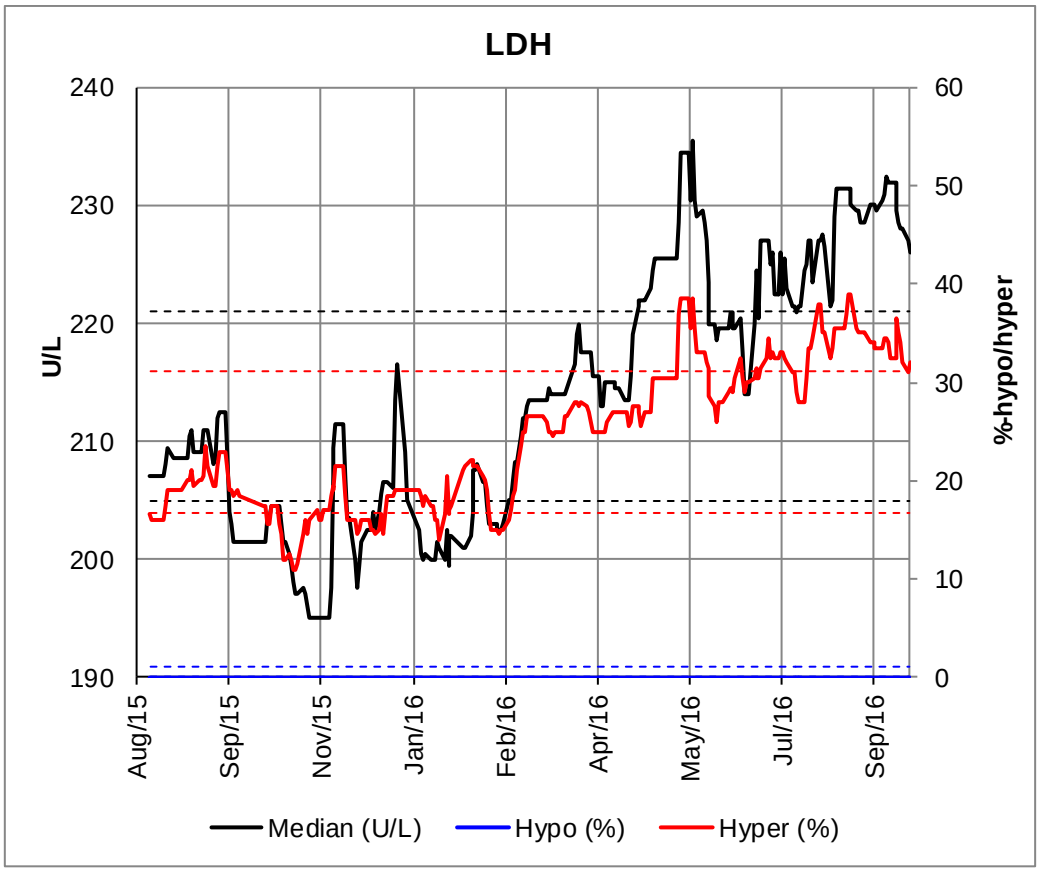
Glucose



The figure shows 3 shifts in glucose values, i.e., a first time from ~5.1 mmol/L to ~5.3 mmol/L, then a shortly lasting shift up to 5.5 mmol/L followed by a return to ~5.4 mmol/L, and finally a third shift up to ~5.6 mmol/L. The hypo flagging rate is little affected and is in the order of 1 to 2%, while the hyper flagging rate mainly increases due to the last upwards shift from ~35% to nearly 55%. Note, the Percentiler limit is 0.15 mmol/L (= 3.1%; “desirable” = 2.3%) and the Flagger limit is 20% of the long-term laboratory median.

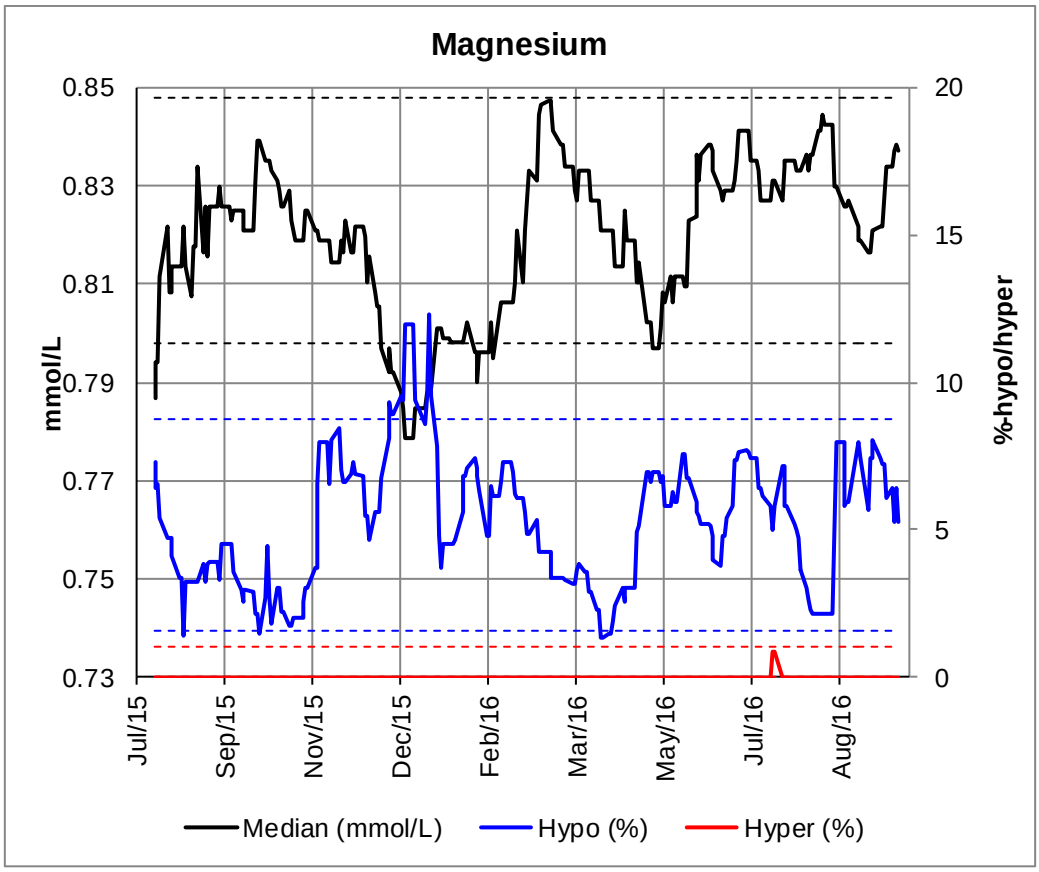
Similar as for total-cholesterol, the glucose example demonstrates the maturity of the test, which, even, requires better stability (desirable stability = 2.3%) because of its lower biological variation as compared to total-cholesterol.

LDH



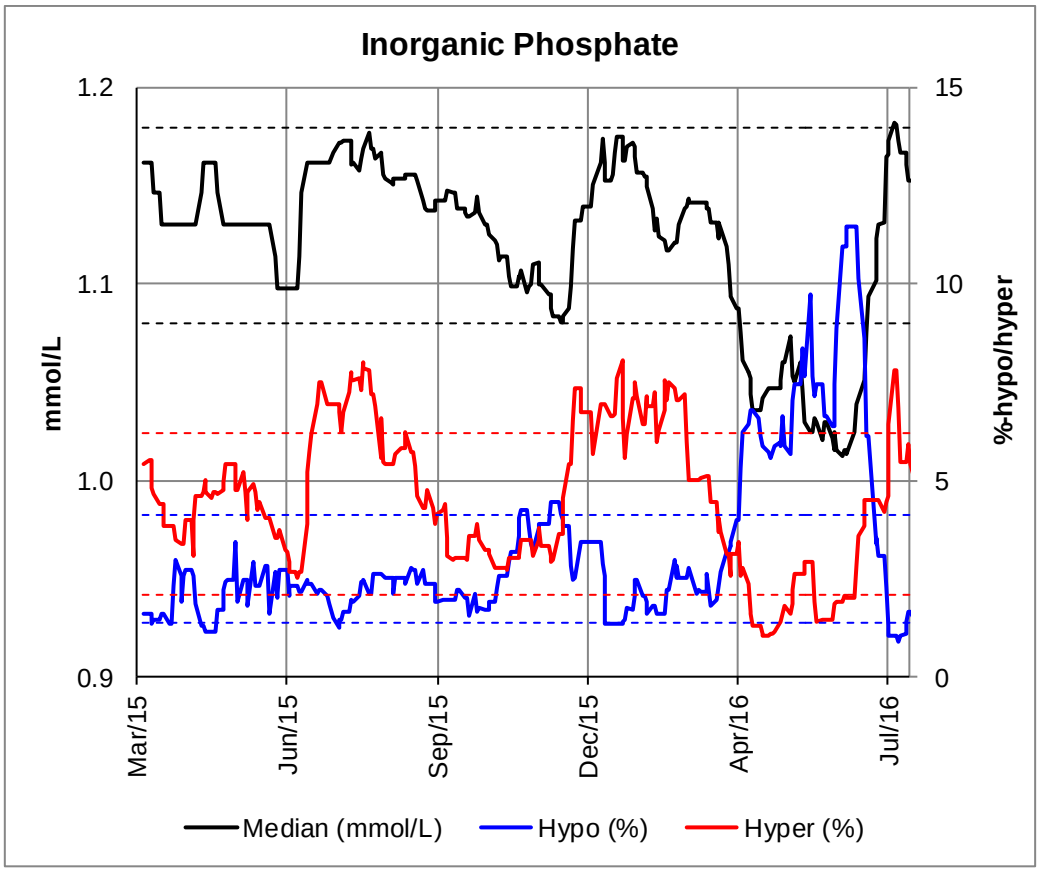
The figure shows drifting LDH values from ~200 U/L up to to ~230 U/L. The hyper flagging rate gradually increases from ~18% to nearly 33%. As is the case for other enzymes, the hypo flagging rate is nearly 0%. Note, the Percentiler limit is 8 U/L (= 4.6%; “desirable” = 4.3%) and the Flagger limit is 30% of the long-term laboratory median.

Magnesium



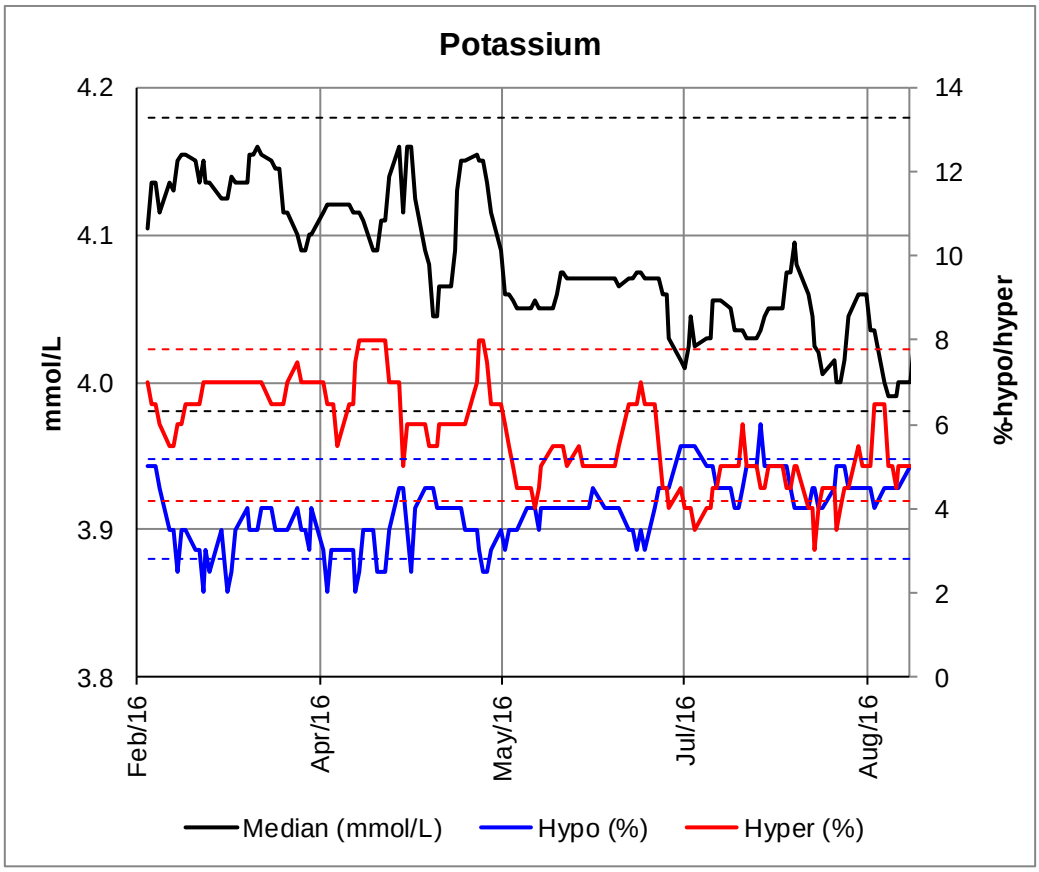
The figure shows several shifts in the magnesium values, i.e., a first drop from ~0.83 mmol/L to ~0.79 mmol/L, followed by an increase to ~0.84 mmol/L, to drop back to ~0.81 mmol/L and finally increase to 0.835 mmol/L. The hypo flagging rate decreases from ~7.5% to ~2.5%, then returns to ~7.5-9.0%, comes down to ~2.5%, to normalize again to ~7.5%. The hyper flagging rate is nearly 0%. Note, the Percentiler limit is 0.02 mmol/L (= 3.0%; “desirable” = 1.8%) and the Flagger limit is 70% of the long-term laboratory median.

Inorganic Phosphate



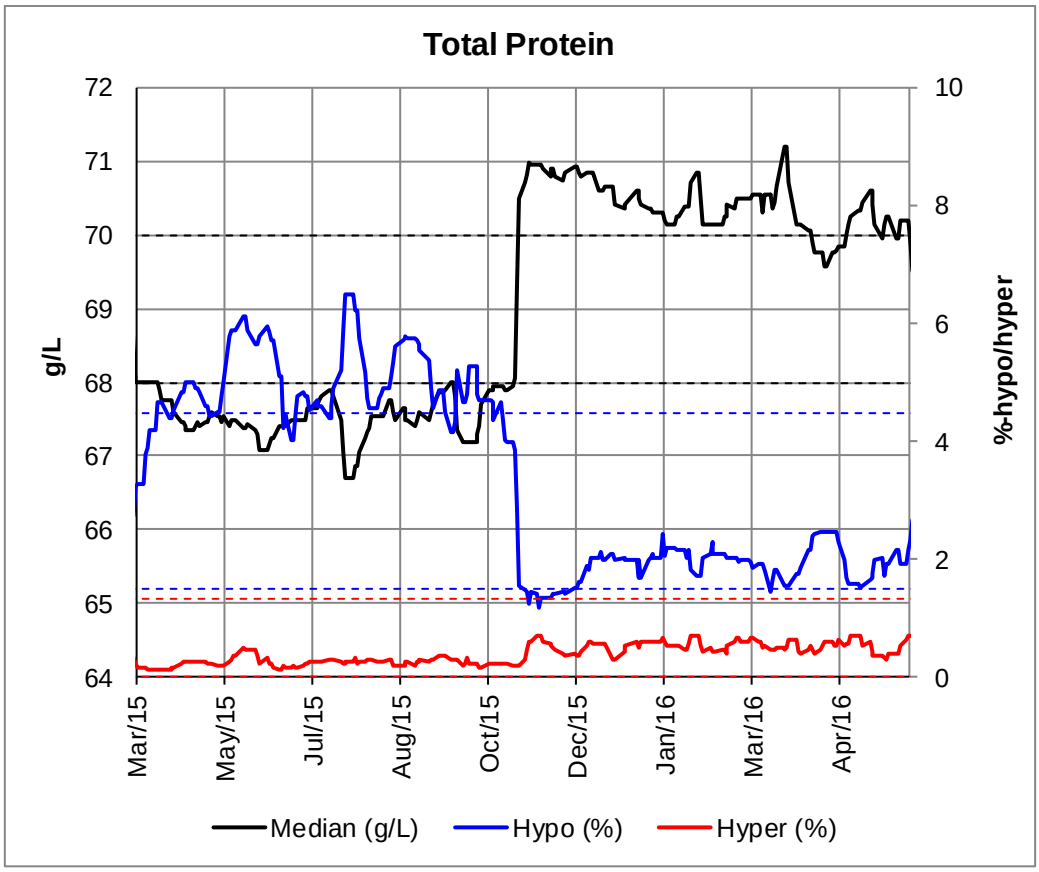
The figure shows several shifts in the inorganic phosphor values, the major ones being from ~1.13 mmol/L to ~1.03 mmol/L and back to ~1.17 mmol/L. The hypo flagging rate concordantly increases from ~2% to ~12% and drops back to ~2%. The hyper flagging rate mainly decreases due to the downwards analytical shift from ~7% to nearly 2.5%. Note, the Percentiler limit is 0.05 mmol/L (= 4.4%; “desirable” = 3.4%) and the Flagger limit is 50% of the long-term laboratory median.

Potassium



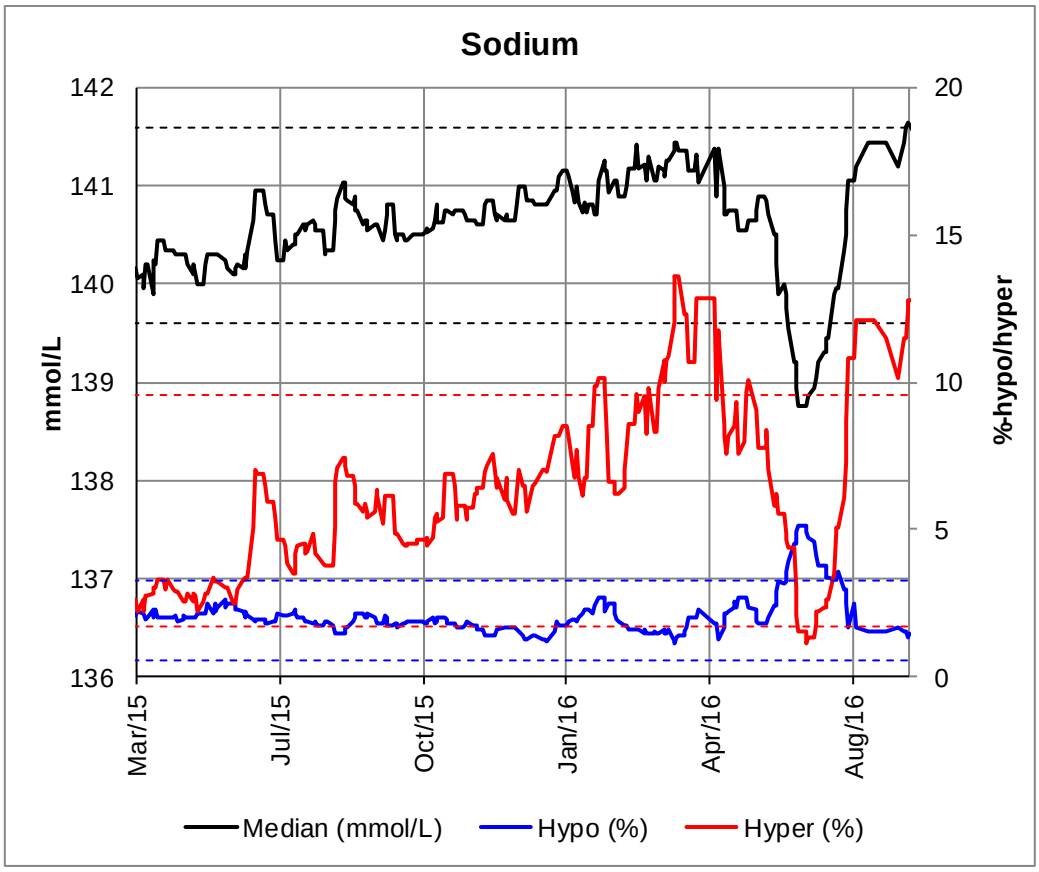
The potassium assay is among the most stable assays across manufacturers and laboratories; typically, the limits are never violated. The figure shows potassium values going down from ~ 4.15 mmol/L to ~4.05 mmol/L. The hypo flagging rate increase concordantly from ~3% to ~5%, while the inverse happens for the hyper flagging rate decreasing from ~7% to nearly 4%. Note, the Percentiler limit is 0.1 mmol/L (= 2.4%; “desirable” = 1.8%) and the Flagger limit is 30% of the long-term laboratory median.

Total Protein



The figure shows a major shift in the total protein values from ~67.5 g/L to ~70.5 g/L. The hypo flagging rate decreases concordantly from ~5% to ~2%. The hyper flagging rate triplicates (from ~0.2% to nearly 0.6%), however, is generally low. Note, the Percentiler limit is 1 g/L (= 1.4%; “desirable” = 1.4%) and the Flagger limit is 50% of the long-term laboratory median.

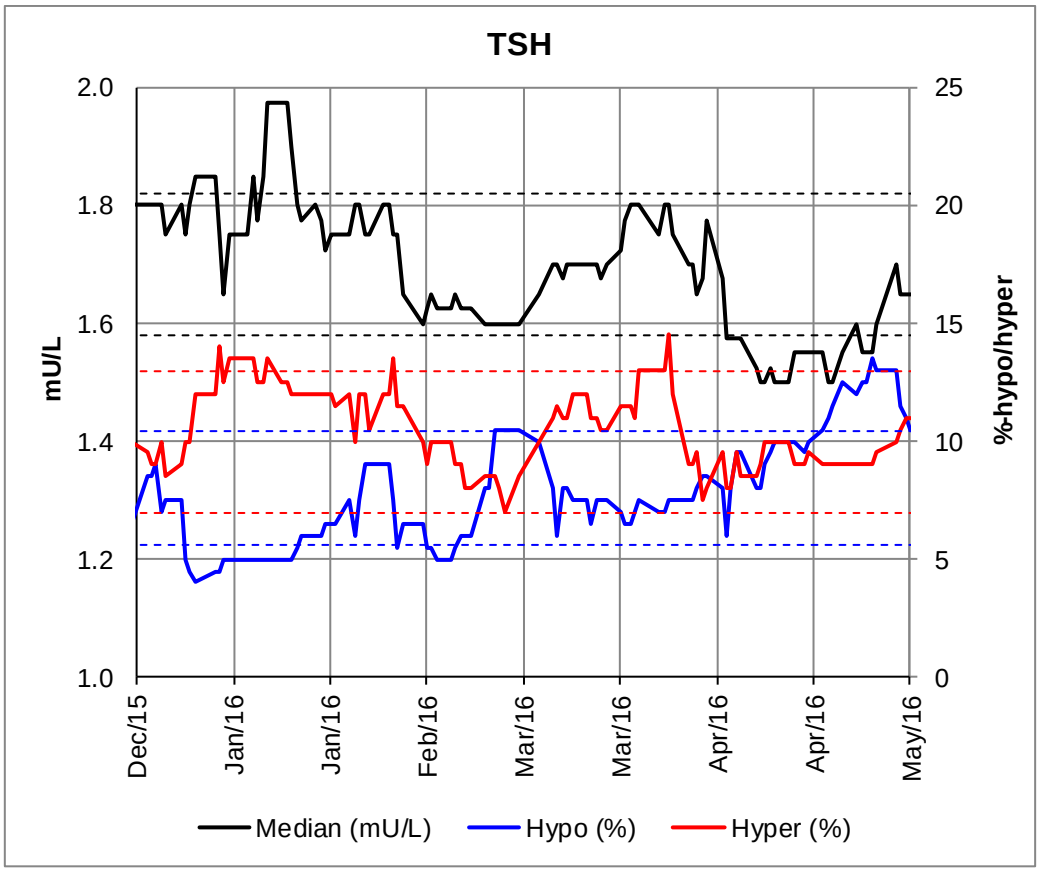
Sodium



The figure shows a drift in sodium values from ~140 mmol/L up to ~141.4 mmol/L, followed by a downwards shift to ~139 mmol/L, and a normalization back to 141.5 mmol/L. The hypo flagging rate is affected by the shift and changes from ~2% to ~5%, and normalizes back to ~2%. The hyper flagging rate is affected by the drift as it increases from ~3% up to ~12%; due to the downward shift it drops from ~12% to nearly 2% and then increases back to ~12%. Note, the Percentiler limit is 1 mmol/L (= 0.7%; “desirable” = 0.2%) and the Flagger limit is 70% of the long-term laboratory median.

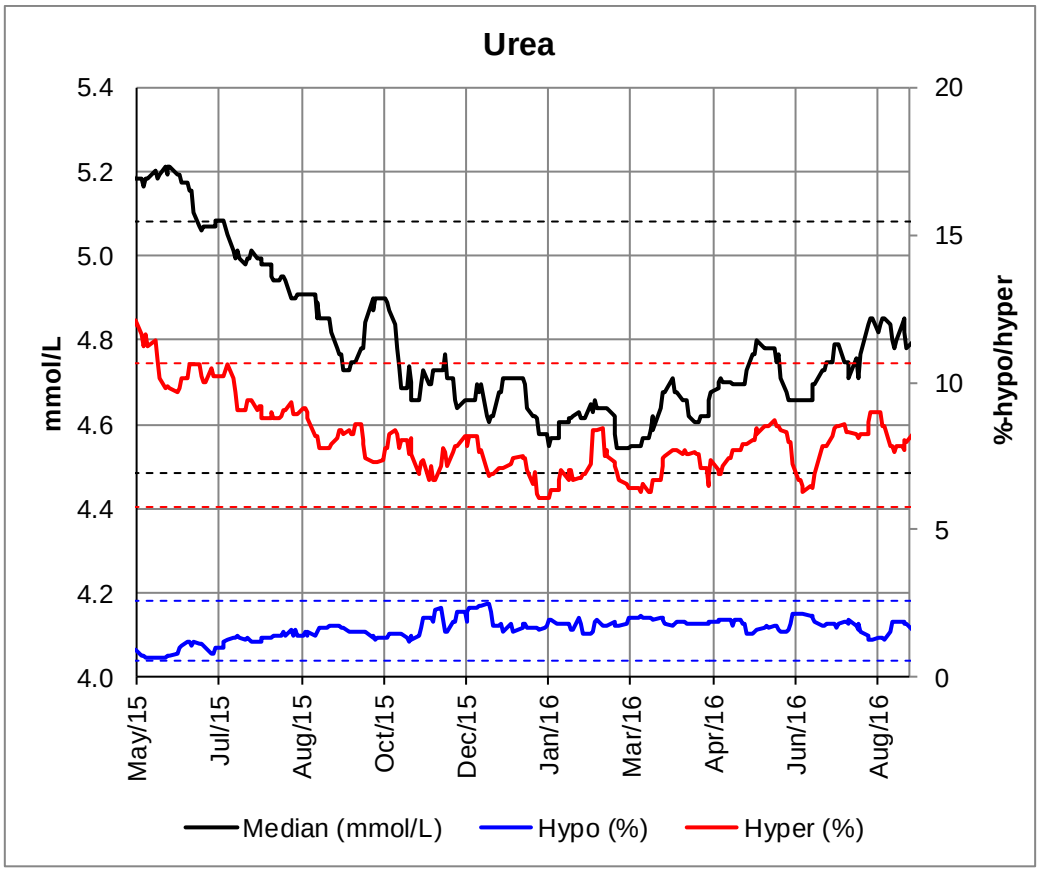
The hyper flagging range nearly triplicated (3% to 9%) due to a drift from 140 to 141 mmol/L. Achieving a 1 mmol/L stability could significantly improve the clinical utility of the sodium test. Indeed, several laboratories were able to reach such a stability during more than 2 years, supporting the applicability of the 1 mmol/L stability which, however, still can result in significant increase in flagging rates.

TSH



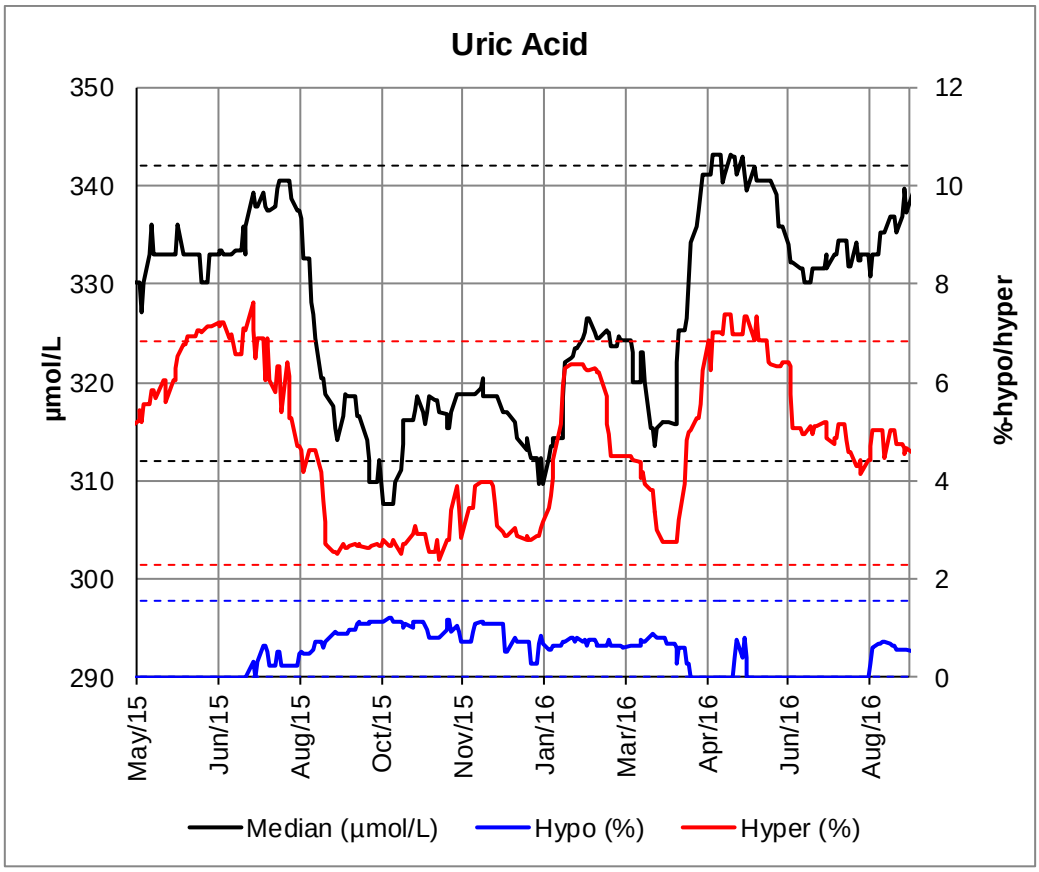
The figure shows variable TSH values ~1.8 mIU/L to ~1.55 mIU/L. The hypo flagging rate follows the trend grossly and varies between ~5% and 12%. The hyer flagging rate follows the analytical trend nicely and varies between ~7% and 13%. Note, the Percentiler limit is 0.12 mmol/L (= 7.7%; “desirable” = 7.8%) and the Flagger limit is 30% of the long-term laboratory median.

Urea



The figure shows drifting urea values from ~5.2 mmol/L to ~4.6 mmol/L and, finally slightly increasing to ~4.8 mmol/L. The hypo flagging rate <1% gradually increases up to ~2.5% and levels off at ~1.5%, while the hyper flagging rate decreases from ~12% to nearly 7.5%. Note, the Percentiler limit is 0.3 mmol/L (= 6.0%; “desirable” = 5.6%) and the Flagger limit is 30% of the long-term laboratory median.

Uric acid



The figure shows a downward and upward shift in the uric acid values from ~335 µmol/L to ~315 µmol/L and back to ~340 µmol/L. Due to the downward analytical shifts, the hyper flagging rate at ~7% falls to ~2.5%; it returns to nearly 7% after the upwards shift. As expected, the hypo flagging rate is very low and little affected by the variation and varies between 0 and 1%. Note, the Percentiler limit is 15 µmol/L (= 4.8%; “desirable” = 4.9%) and the Flagger limit is 50% of the long-term laboratory median.

Conclusion

Notwithstanding the benefits from participating in the combined Percentiler/Flagger application, there are still limitations to mention. One is that the observations in the Flagger application are of limited value for low-throughput laboratories because of high variability of the flagging rates that cannot be compensated for by choosing a higher n for calculating the moving median. However, we believe that even these laboratories can profit from the application simply by being part of it and learning from the other participants via the reports we send. Another limitation is that currently only monitoring the stability of the flagging rate at the individual laboratory level is possible. The number of participants is still too low to do the comparison across laboratories or peer groups. But even if this would be possible, it would be meaningless as long as the current design of the Flagger does not give insight in the locally applied decision limits for flagging of results. Indeed, deviations from the peer potentially result from the use of different reference interval limits. Therefore, it is our plan for the near future to compare the locally used cut-off points. Either, this may clarify the reasons why different cut-offs are justified, or, if none can be identified, this may become an opportunity to harmonize them.

Overall, we showed that the combination of monitoring flagging rates together with daily patient medians enables to translate the effect of analytical variation on surrogate medical decisions. Our observations re-iterated the utility of the concept of setting analytical performance specifications from biological variation. The Flagger/Percentiler application can form a bridge of the “medium-level” hierarchy for setting analytical specifications (biology) to the “top-level” (clinical situations). The advantage of the application is the direct visualization of analytical instability and its effect on flagging of results at the individual participant level using data readily available in the laboratory itself.