Linda M. Thienpont* and Dietmar Stöckl

**Percentiler and Flagger – low-cost, on-line monitoring of laboratory and manufacturer data and significant surplus to current external quality assessment**

https://doi.org/10.1515/labmed-2018-0030
Received March 10, 2018; accepted April 27, 2018; previously published online June 14, 2018

**Abstract**

**Background:** We developed two web-based applications called the “Percentiler” and “Flagger”. They use electronically sent data from the analysis of patient samples (medians in the Percentiler; % flagging in the Flagger). Through a graphical user interface, the applications allow on-line monitoring of the stability of analytical performance and flagging rate, both assessed against quality specifications. These are guided by biological variation (Percentiler) and effect of analytical instability on surrogate medical decisions (Flagger). Here, we report on the use of the applications.

**Methods:** We constructed examples with combined observations to investigate whether the Flagger adequately translates the effect of analytical instability observed in the Percentiler, and whether the changes in the flagging rate tolerated by the proposed stability limits is realistic in combination with the analytical performance goals.

**Results:** In general, the examples show that the most prominent flagging rates correlate well with the analytical stability and that the limits proposed for the Flagger are realistically linked to those of the Percentiler. They also show that for certain analytes the specifications for stable flagging rates can be restricted to 20% (relatively to the laboratory’s long-term flagging median) despite ambitious analytical performance goals, while for others they need to be expanded up to 70% in concordance with decreasing biological variation.

**Conclusions:** The examples confirm that the changes in flagging rate is well related to the analytical variation, and that the proposed stability limits are fit-for-purpose. The combined observations may help individual laboratories to define realistic but ambitious performance specifications that apply for their local situation.

**Keywords:** biological variation; hypo- and hyper-flagging rate; moving median; performance specifications; stability of performance.

**Brief summary:** The Percentiler and Flagger use data from the analysis of patient samples (medians, % flagging against local cut-off points). Through a graphical user interface, participants can monitor online the stability of their performance and the impact of analytical variation on surrogate medical decisions, both assessed against stability limits. Experience from combining both applications helps in defining realistic but ambitious performance goals.

**Introduction**

Laboratories are aware that providing reliable test data produced under stable performance conditions is a prerequisite for clinicians for proper diagnosis and monitoring of patients or therapies. As of today, laboratories mostly utilize commercial test systems/assays and they rely on the intrinsic quality claimed by the in-vitro diagnostic (IVD) industry. Intrinsic quality refers to common analytical attributes like precision, accuracy, specificity, etc., however, nowadays, the focus is also on metrological traceability and/or comparability across manufacturers [1]. Of course, after implementation of test systems laboratories have to sustain the intrinsic quality and stability of performance. For this purpose, they use quality assurance systems comprising internal quality control (IQC) and external quality assessment (EQA). However, as indispensable these measures may be, it should not be underestimated that conventional IQC and EQA are plagued by systematic issues and/or limitations. A major problem is related to the question of what constitutes desirable quality and stability. Despite the development of several scientific concepts to derive specifications, the profession continues to struggle with implementing generally accepted “numbers” [2, 3]. The only consensus reached is on classifying the concepts in a three-level hierarchy: the clinical outcomes concept at the top, the one based on...
biological variation in the middle and the state-of-the-art at the bottom [4]. Another key issue is due to the sample materials used, which are mostly "processed" (e.g. pooled, stripped, dialyzed, etc.). Admitted, these kinds of materials have desirable advantages for use as controls: they are cheap, available in high volumes, easy to supplement to obtain pathological levels, can be stabilized by lyophilization, etc. However, the problem is that they do not necessarily reflect the reality of patient testing. In metrological terms, one speaks of the potential "non-commutability" of a material with routine assays [5]. In IQC, non-commutable materials may miss problems in the analytical phase, such as trends or shifts due to reagent or calibration lot changes [6]. In EQA, they may point to a bias which does not exist in patient samples. The latter makes them counterproductive to demonstrate comparability across systems [7]. Therefore, conventional EQA must necessarily restrict to the peer group level. However, peers should ideally be homogeneous, i.e. from grouping by test system combined with reagent, calibrator and assay from the same IVD manufacturer, and not by method principle. This is inadequate because the individual design and optimization of test systems using the same principle may make them non-equivalent. The best current solution to proper peer group building is provided by commercial programs for combined IQC/EQA. But still these programs have the limitation that the data are mostly not accessible in real-time but, for example, only on a monthly basis. A similar problem is faced in regulatory EQA schemes conducted at low frequency. As a result, laboratories become aware of problems in performance compared to their peer only after the event. In addition, most providers of combined IQC/EQA programs do not critically report on performance data of individual IVD manufacturers, but leave the interpretation to the participants. This practice stems from the commercial surroundings in which they operate, which hampers open communication. Theoretically, regulatory EQA organizers are in the position to do so, but as mentioned before, they often cannot, hampered as they are by the lack of proper peer groups. As a result, both commercial and regulatory EQA schemes miss the basis to serve as a quality-improvement tool.

In an attempt to respond to the above limitations by an alternative way of quality assessment, we developed two web-based applications called the “Percentiler” and “Flagger” [8–12]. By design, both are based on establishing databases from analysis of patient samples, which are per definition commutable with routine test systems. The applications are up to now freely available (free-of-charge) to interested laboratories. All they have to do is establish IT connectivity with our MySQL databases to send their data. We provide them with a graphical user interface for on-line monitoring of the stability of their analytical performance compared to their peers (Percentiler) and of the impact of instability on the flagging rate (Flagger). Hence, the applications can serve the purpose of complementary IQC from the analysis of patient samples, while covering long observation time; in addition, they are a low-cost significant surplus to EQA: instead of using extra samples, the databases are from readily available results and directly allow peer group comparisons across manufacturers; these comparisons provide information at the relevant level (patient samples) and in a timely manner to remediate. Indeed, we oversee all data, and, as independent operators without any external funding or sponsoring, we can openly communicate. In contrast, we have a written or silent confidentiality agreement with laboratories; however, they themselves can share their data with IVD manufacturers, if they wish. Altogether, it is our aim to act via the applications as a mediator between laboratories and IVD manufacturers, so that they become empowered in their common interest to improve analytical quality to optimum patient care.

Here, we report on the link between the Percentiler and Flagger applications.

**Materials and methods**

The design of the Percentiler and [Flagger Thienpont & Stöckl Wissenschaftliches Consulting GbR, Rennertshofen (OT Bertoldsheim), Germany] is described elsewhere [9, 10, 12]; however, to facilitate understanding, we repeat the key points. The applications are not commercially exploited by the developers. Both use databases established from instrument-specific medians calculated from outpatients’ results (Percentiler) and the number of results (in % relative to the total number) flagged when they exceed the locally used cut-off points (Flagger). Preferably, the calculations are done by an automatic function either in the laboratory information system (LIS) or via a middleware or home-made solution. The same applies for electronic data transmission preferably on a daily basis, or if not possible, batch-wise (frequency as convenient). Note that some LIS providers offer to their customers a free-of-charge solution for these requirements, others do it at a low cost.

Initially, we focused on 20 clinical chemistry analytes and two thyroid hormones [9, 11]. Recently, we expanded the analytes based on demand of participants (additional lipids, vitamin B12, folate, ferritin, immunoglobulins,
glycated hemoglobin, and the most common hematological parameters).

The Percentiler and Flagger user interfaces are accessible on-line at https://www.thepercentiler.be/ and https://www.theflagger.be/, respectively (more details in Refs. 9, 10). Authorization is secured via a laboratory-specific username and password (note, organizations with multiple sites receive a group-specific password). A demo version of the interfaces can be accessed via the account “DEMOLAB” (=username) and “demo1234” (=password). Whereas we operators have access to the complete database, participating laboratories are restricted to their own data. Currently, ~135 laboratories participate in the Percentiler with double the number of instruments (geographic distribution shown in Supplementary Figures 1S and 2S); approximately half of them also use the Flagger application. Whereas some laboratories have already participated since 2014, new laboratories regularly join.

The functionalities of the Percentiler and Flagger user interfaces are very similar. They allow the participating laboratory to plot and download for each analyte the instrument-specific time course of the moving medians (Percentiler; examples shown in Refs. [9–11]) and of the % hypo- and hyper-flagging (Flagger; for an example see the Supplementary Figure 3S). In addition, the long-term median and the all or peer group moving median is shown. The charts also include stability limits delineating a shaded zone on both sides of the laboratory’s long-term flagging median to indicate at a glance whether the performance and/or flagging rates are stable on the mid- to long-term timescale (see the Supplementary Figure 3S). Note that the Percentiler stability limits are guided first by biological variation and, if not feasible, by the state-of-the-art performance [9, 11]. For the Flagger, they are based on an earlier developed concept that investigates the effect of a bias (in fractions of the biological variation) on surrogate medical decisions [13, 14]. For the latter, we used flagging of results exceeding locally used cut-offs (mostly the 1.96σ decision point for Gaussian distributed reference interval data, which is typically the “first alert level” that triggers physicians). When we found that the bias specification according to the Gowans et al. concept

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Flagger specifications, %</th>
<th>Percentiler specifications, % (absolute)</th>
<th>Specifications for desirable bias, % [16]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHOL</td>
<td>20</td>
<td>3.8 (0.2 mmol/L)</td>
<td>4.1</td>
</tr>
<tr>
<td>GLUC</td>
<td>20</td>
<td>3.1 (0.15 mmol/L)</td>
<td>2.3</td>
</tr>
<tr>
<td>ALP</td>
<td>30</td>
<td>6.8 (5 U/L)</td>
<td>6.7</td>
</tr>
<tr>
<td>ALT</td>
<td>30</td>
<td>9.5 (2 U/L)</td>
<td>11.5</td>
</tr>
<tr>
<td>AST</td>
<td>30</td>
<td>6.5 (1.5 U/L)</td>
<td>6.5</td>
</tr>
<tr>
<td>BIL</td>
<td>30</td>
<td>10.0 (1 μmol/L)</td>
<td>9.0</td>
</tr>
<tr>
<td>CREAT</td>
<td>30</td>
<td>3.9 (μmol/L)</td>
<td>4.0</td>
</tr>
<tr>
<td>CRP</td>
<td>30</td>
<td>9.6 (0.25 mg/L)</td>
<td>21.8</td>
</tr>
<tr>
<td>FT4</td>
<td>30</td>
<td>3.3 (0.5 pmol/L)</td>
<td>3.3</td>
</tr>
<tr>
<td>GGT</td>
<td>30</td>
<td>9.1 (2 U/L)</td>
<td>11.1</td>
</tr>
<tr>
<td>K</td>
<td>30</td>
<td>2.4 (0.1 mmol/L)</td>
<td>1.8</td>
</tr>
<tr>
<td>LDH</td>
<td>30</td>
<td>4.6 (8 U/L)</td>
<td>4.3</td>
</tr>
<tr>
<td>TSH</td>
<td>30</td>
<td>7.7 (0.12 mU/L)</td>
<td>7.8</td>
</tr>
<tr>
<td>UREA</td>
<td>30</td>
<td>6.0 (0.3 mmol/L)</td>
<td>5.6</td>
</tr>
<tr>
<td>ALB</td>
<td>50</td>
<td>2.3 (1 g/L)</td>
<td>1.4</td>
</tr>
<tr>
<td>CL</td>
<td>50</td>
<td>1.0 (1 mmol/L)</td>
<td>0.5</td>
</tr>
<tr>
<td>PHOSP</td>
<td>50</td>
<td>4.4 (0.05 mmol/L)</td>
<td>3.4</td>
</tr>
<tr>
<td>PROT</td>
<td>50</td>
<td>1.4 (1 g/L)</td>
<td>1.4</td>
</tr>
<tr>
<td>UA</td>
<td>50</td>
<td>4.8 (15 μmol/L)</td>
<td>4.9</td>
</tr>
<tr>
<td>CA</td>
<td>70</td>
<td>1.7 (0.04 mmol/L)</td>
<td>0.8</td>
</tr>
<tr>
<td>MG</td>
<td>70</td>
<td>3.0 (0.02 mmol/L)</td>
<td>1.8</td>
</tr>
<tr>
<td>NA</td>
<td>70</td>
<td>0.7 (1 mmol/L)</td>
<td>0.2</td>
</tr>
</tbody>
</table>

They are reflected against those used in the Percentiler, as well as the specifications for desirable bias inferred from the biological variation concept (note that we used the database cited in Ref. [16]). The table is sorted by increasing specifications in the Flagger; the absolute minimum stability limit in the Flagger is set to 1%. *Abbreviations for analytes (in alphabetical order): ALB, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BIL, total-bilirubin; CA, calcium; CHOL, total-cholesterol; CL, chloride; CRP, C-reactive protein; CREAT, creatinine; FT4, free thyroxine; GGT, γ-glutamyltransferase; GLU, glucose; K, potassium; LDH, lactate dehydrogenase; MG, magnesium; NA, sodium; PHOSP, inorganic phosphate; PROT, total-protein; TSH, thyroid-stimulating hormone; UA, uric acid.
of common reference intervals (0.25*CV group) causes an increase of false positives by ~80% [15], we proposed to set the allowable increase to 30% [14].

We selected five laboratories to construct examples of observations made in both applications. We used them to answer the questions whether the Flagger adequately translates the effect of the analytical variation/instability observed in the Percentiler, and whether the changes in flagging rate tolerated by the proposed stability limits is realistic in combination with the analytical performance goals.

Results

Table 1 shows the Flagger and Percentiler specifications we used for the 22 examined analytes. These are limits that should not be violated during more than 1 week. This timeframe was chosen to prevent “over” interpretation of the performance. Note that the Flagger specifications are semi-arbitrary: we started with applying 30% as proposed by Stepman et al. [14] and found that many analytes were in agreement with the current state-of-the-art performance at the concentration levels of the reference interval percentiles. However, sometimes we had reasons to use more stringent or looser specifications (ranges between 20% and 70%): for total-cholesterol and glucose we defined the most stringent specification (20%), in view of their paramount importance in public health policies (coronary artery disease and diabetes); for eight other analytes, we used looser specifications (50% and 70%), in particular, for those under tight physiological control (e.g. calcium, magnesium, sodium; see also Discussion). Note that the Flagger specifications are applied relatively to the long-term flagging median observed for the individual laboratory. For example, if a laboratory has a long-term flagging median of 10% for aspartate aminotransferase (AST), then the limit is set at ±3% (=30% of 10%); however, in case it is <3.3%, it is set to a minimum value of ±1%.

Below, we present our combined observations for the five selected examples. The Supplementary Figures 4S to 25S show examples for all 22 analytes.

Cholesterol

Figure 1 shows a major downwards shift in total-cholesterol median values from ~4.7 mmol/L to ~4.4 mmol/L. The hypo-flagging rate is nearly unaffected and is on the order of ~1%, while the hyper-flagging rate decreases from ~36% to nearly 23%. The figure shows in addition that operating the test with an analytical stability within the ±0.2 mmol/L Percentiler limit would result in hyper-flagging rates between 24 and 37%.

Glucose

Figure 2 shows three shifts in glucose values, a first shift 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. Mainly the last upwards shift causes the hyper-flagging rate to increase from ~35% to nearly 55%, while the hypo-flagging rate is little affected and is on the order of 1% to 2%. In addition, the figure demonstrates that operating the test with a stability within the ±0.15 mmol/L Percentiler limit would result in hyper-flagging rates between 33 and 50%.

ALT

Figure 3 shows a drift of the alanine aminotransferase (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 then drops to 0%, while the hyper-flagging rate decreases from ~7% to nearly 0% and then rises.
In addition, the figure shows that operating the test within the ±2 U/L Percentiler limit would result in hypo-flagging rates between 9 and 17%. Note that we found the ALT tests stable within the selected Flagger limit of 30% (data not shown) and that most laboratories only have hyper-flagging limits for ALT and the other monitored enzymes.

Calcium

Figure 4 shows moderately varying calcium values, i.e. they first shift upwards from ~2.33 mmol/L to ~2.38 mmol/L, then fall back to 2.33 mmol/L, followed by a second gradual increase to ~2.42 mmol/L, to finally drop to 2.30 mmol/L. The analytical shifts most strongly affect the hypo-flagging rate, which decreases from ~13% to ~3%, then increases to ~9%, falls back to 4% to finally increase a second time to ~15%. No effects are observed for the hyper flagging due to its very low rate. In addition, the figure demonstrates that operating the test with a stability within the ±0.04 mmol/L Percentiler limit would result in hypo-flagging rates between 2 and 10%.

Sodium

Figure 5 shows a drift in the sodium values from ~140 mmol/L up to ~141.4 mmol/L, followed by a downward shift to ~139 mmol/L, and a normalization back to again to ~8%. In addition, the figure shows that operating the test within the ±2 U/L Percentiler limit would result in hypo-flagging rates between 9 and 17%. Note that we found the ALT tests stable within the selected Flagger limit of 30% (data not shown) and that most laboratories only have hyper-flagging limits for ALT and the other monitored enzymes.
141.5 mmol/L. The hypo-flagging rate is most affected by the shift, which causes it to change from ~2% to ~5%, and then to normalize back to ~2%. The hyper-flagging rate is most influenced 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%. In addition, the figure shows that operating the test within the ±1 mmol/L Percentiler limit would result in hyper-flagging rates between 2 and 12%.

**Discussion**

Our long-term experience with the Flagger and Percentiler allowed us to investigate in more detail the link between the two applications. The emphasis was, in addition, on the link between the specifications we proposed for the stability of surrogate medical decisions and analytical performance. Note that we used the flagging rate against locally used cut-offs (mostly the 2.5th and 97.5th percentiles of the reference interval) as surrogate medical decision. As the Percentiler monitors data with concentrations corresponding to the 50th percentile of a reference interval, one may argue that the observed analytical variation cannot be representative of changes in flagging. However, the majority of the selected examples waived this argument as the most prominent flagging rates (be it at the 2.5th or 97.5th centiles) correlate quite well with the Percentiler observations on instability.

We demonstrate here that restricting the changes in flagging rate to 20% of the long-term flagging median is realistic for some clinically very important analytes (total-cholesterol and glucose), despite the fact that their relatively low biological variation requires ambitious analytical stability limits (for example, 2.3% for glucose; see Table 1). This is in contrast with the inorganic phosphate case (with an analytical stability limit of 3.4%), where we have to operate the Flagger with a 50% limit. We speculate that this is because manufacturers (and laboratories) give special attention to the testing of analytes that are in the public or scientific focus, which does not apply for the inorganic phosphate test. We can apply a 30% stability limit in the Flagger for most other analytes, for which the limits used in the Percentiler correspond with the bias derived according to the biological variation concept. In contrast, for the tightly controlled analytes (albumin, chloride, total-protein, calcium, magnesium, sodium) the Flagger limits need to be broader (50% or 70%). This is concordant with the need to extend the Percentiler bias specifications beyond those inferred from biological variation [9, 10]. Indeed, the current state-of-the-art analytical performance for these analytes is not appropriate to strictly apply this concept, e.g. for sodium, we use a bias limit of 1 mmol/L instead of 0.32 mmol/L according to the biological variation concept (0.32 mmol/L is 0.23% at a median concentration of 140.5 mmol/L). We also have to apply a 50% limit for uric acid because of its seasonal variation (somewhat higher in the summer). As such, the selected examples give an insight into the effect of analytical variation on local flagging rates. Note that our current choice of Flagger specifications is not binding but can be adapted after consultation with its users.

The selected examples also demonstrate in more detail the utility of combining the observations from the Flagger and Percentiler. We believe that doing this can strengthen the awareness of the profession on the importance of quality specifications and can guide them to infer realistic goals that are helpful to improve the clinical utility of laboratory testing. For example, the total-cholesterol case (Figure 1) demonstrates the analytical maturity of the test: we see that several laboratories nicely can keep the flagging rate within the 20% limit during >2 years, which corresponds with working at an analytical stability within 0.2 mmol/L (desirable stability = 3.8%). Similar holds true for the glucose test which even requires better analytical stability (desirable stability = 2.3%) (Figure 2). The ALT tests typically are stable within the proposed limits (not shown), except the one used to construct (Figure 3). The
example demonstrates that the analytical instability of the test gives rise to significant changes in flagging rates (the hypo flagging varies from 0% to 28%, the hyper from 0% to 8%). This is highly contra-productive for the utility of the ALT test for newer applications such as early detection and/or monitoring of metabolic changes (“metabolic syndrome”) which require an analytical stability in the order of 2 IU at concentrations in the reference interval [17–19]. In that connection, establishing cut-offs for high-normal ALT values may be useful as currently most laboratories only flag very high results. As reported earlier, calcium flagging rates are very much influenced by analytical instability because of the low biological variation [12]. The hypo flagging rate is ~15% at 2.30 mmol/L, whereas it is only ~3% at 2.42 mmol/L in the example shown in Figure 4. This means that operating the calcium test with an instrument with analytical bias of ~0.12 mmol/L can result in five-fold different hypo flagging rates. Also, the sodium example in Figure 5 excellently demonstrates the importance of linking the observations on flagging and analytical stability. It confirms our earlier observations that the hyper flagging range nearly triplicates (3% to 9%) due to a drift in the moving median from 140 to 141 mmol/L [20]. Note that this drift is entirely within the allowable performance variation we proposed and found feasible for the Percentiler, as several laboratories succeed in sustaining it for more than 2 years. However, if we interpret the example the other way around, we have to conclude that, even when running a sodium test within the proposed stability limits, the flagging rate is very much impacted. This has to be taken into account with regard to the clinical utility of the sodium test [21, 22].

Notwithstanding the benefits of the combined applications, there are some limitations to mention. First, the observations in the Flagger application are of limited value for low-throughput laboratories because of high variability in 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 a part of it and learning from the other participants via the reports we regularly 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 make 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 an insight into the locally applied decision limits for flagging. Indeed, deviations from the peer potentially result from the use of different cut-off points. Therefore, it is our plan for the near future to compare the used decision points. Either, this may clarify why different cut-offs are justified or it may become an opportunity to harmonize them. Last but not the least, we have to admit that the statement in the title and introduction that the Percentiler is to be seen as a significant surplus to the current EQA will only become a reality once we have sufficient laboratories on board who use the most important commercial test systems/assays. According to our previous experience with EQA, we estimated that around 20–25 laboratories are needed to provide meaningful conclusions on the performance of a homogeneous peer group [23]. For some peers, we have reached that point already, while for others, we still need substantiation. This may be achievable if one or more EQA organizers would be interested to host the Percentiler as surplus to their classical schemes. Of course, it is not to be excluded that a fee will be charged; however, it will probably be reasonably low as it only should cover the costs of developing adequate tools for data treatment, without any need to purchase/dispatch extra samples.

Conclusions

Overall, this report shows that it is a big asset to use data from analysis of patient samples to visualize different aspects of analytical stability in laboratory testing. The constructed examples particularly show that the Flagger adequately translates the effect of analytical variation on surrogate medical decisions. The examples re-iterate the utility of setting analytical performance specifications from the biological variation concept. In addition, they show that combining the observations from the Flagger and Percentiler makes it possible to bridge the “medium-level” hierarchy for setting analytical specifications (biological variation) to the “top-level” (clinical outcomes). Last but not least, the observations may help laboratories to define realistic yet ambitious performance goals that apply for their local situation.

Acknowledgments: The authors are indebted to the participating laboratories. They appreciate their interest in the applications and the long-term reporting of results. The authors also express their gratitude to Bruno Neckebroek for his assistance in standardizing the data and developing a Java-based application for the data processing and user interface.

Author contributions: Both authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.
Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The IVD diagnostic industry played no role in the study design, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References


Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/labmed-2018-0030).