Anne Elisabeth Solsvik*, Ann Helen Kristoffersen, Sverre Sandberg, Gro Gidske, Anne Vegard Stavelin, Joakim Eikeland and Erik Amundsen

A national surveillance program for evaluating new reagent lots in medical laboratories

https://doi.org/10.1515/cclm-2021-1262
Received December 2, 2021; accepted January 10, 2022; published online January 19, 2022

Abstract

Objectives: Differences between laboratory results attributable to the use of different reagent lots can potentially affect the diagnosis and monitoring of patients. To minimize patient risks, all laboratories should verify that new reagent lots meet agreed analytical performance specifications (APS). We propose a simplified, pragmatic approach for laboratories that involves compilating results into a national surveillance program, and present the first results obtained when applying this approach to troponins, glycated hemoglobin (HbA_{1c}), prostate-specific antigen (PSA) and D-dimer.

Methods: In the surveillance program we have (i) determined APS for selected analytes, (ii) implemented a simplified procedure for lot evaluation with patient samples used in laboratories across Norway and (iii) performed central processing of the results from the participating laboratories.

Results: Over a one-year period, 27 Norwegian laboratories returned results from 28 lot changes for troponin I, 11 for troponin T, and 29 for HbA_{1c}, PSA and D-dimer. The mean difference between two reagent lots was 4.5% for troponin I (for a concentration interval of 20–32 ng/L), 5.1% for troponin T (10.7–17.5 ng/L), 2.2% for HbA_{1c} (40–50 mmol/mol), 3.7% for PSA (3–5 µg/L) and 5.5% for D-dimer (0.4–1.0 mg/L FEU).

Conclusions: A novel procedure for reagent lot evaluation is proposed in which information about multiple lot changes from different medical laboratories can be accumulated nationally. Sharing this information allows simplification of lot evaluations in individual laboratories and provides real-world data about lot-to-lot variations.

Keywords: lot changes; national surveillance; reagent lot-to-lot variation.

***Corresponding author: Anne Elisabeth Solsvik**, MSc, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Noklus, Box 6165, 5892 Bergen, Norway, E-mail: anne.elisabeth.solsvik@noklus.no
Ann Helen Kristoffersen, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway
Sverre Sandberg, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; and Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen, Bergen, Norway
Gro Gidske, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen, Bergen, Norway
Anne Vegard Stavelin, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway
Joakim Eikeland, Department of Medical Biochemistry, Oslo University Hospital, Oslo, Norway, E-mail: joaeik@ous-hf.no
Erik Amundsen, Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Medical Biochemistry, Oslo University Hospital, Oslo, Norway

Introduction

Clinically important differences in measured analyte concentrations after changing a reagent lot are a significant risk that potentially affects patient care [1–4]. It is important for patient results to be consistent over time so that clinicians can accurately compare and interpret new results against previous results, clinical decision limits and reference intervals. Even though the performance of a reagent is validated by its manufacturer, a laboratory needs to ensure that a new reagent lot meets the laboratory's own analytical performance specifications (APS) [5, 6]. However, there are very few publications describing procedures for evaluating lot-to-lot variations in medical laboratories. There are marked variations in both the size and competence of laboratories, and not all laboratories have procedures in place to test for reagent lot-to-lot variations beyond performing internal analytical quality controls. In our experience, laboratory procedures are affected by the

experience with lot comparisons for different measurement procedures (MPs), the significance of such changes for interpreting analytical results and the resources available in the laboratory, and both personnel and financial aspects (e.g., reagent costs).

The Clinical and Laboratory Standard Institute (CLSI) has published the EP26-A guideline for evaluating new reagent lots, entitled "User Evaluation of Between Reagent Lot Variation" [5]. One important disadvantage of the EP26-A guideline is that each laboratory must decide for each analyte the "acceptable critical difference value", the statistical power, and at how many levels and at which concentrations it is clinically relevant to evaluate lot-to-lot variations. Furthermore, the guideline does not describe how to detect a cumulative bias between multiple reagent lots over time.

The EP26-A guideline states that for laboratories with multiple measuring systems (MSs), it is sufficient to evaluate a new lot using only one of the available instruments. Thus, we hypothesized that multiple laboratories in different organizations could share data in order to reduce the workload of each individual laboratory. A more detailed and harmonized procedure for lot verification and a centralized surveillance program would make it possible to collect information about real-world lot-to-lot variations in medical laboratories for different MPs and analytes. Immunological methods are expected to have higher lot-to-lot variation than other clinical chemistry methods [7–9], and so should probably be prioritized when selecting the analytes to use when initially start testing for lot-to-lot variation.

The aims of this paper were to (i) determine APS for reagent lot changes for five selected analytes (immunological methods), (ii) suggest a simplified pragmatic procedure for evaluating new reagent lots and propose how to accumulate consecutive new reagent lot results for the same instrument for each laboratory, and (iii) present the first results from nationally performed processing of the lot change results from the participating laboratories.

Materials and methods

In 2018, the Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus) hosted a workshop on evaluating new reagent lots for medical laboratories. As a follow-up, Noklus initiated a project and established a working group (WG) comprising medical specialists in laboratory medicine and biomedical laboratory scientists, with the purpose of developing a practical procedure for evaluating new reagent lots that could also be used for surveillance of lot-to-lot variations in Norway. This proposal was based upon experience from the workshop, a systematic literature review, information on how reagent lot-to-lot variations have been evaluated at the university hospitals of the WG members and responses to a short questionnaire sent to medical laboratories. Five analyte-specific Excel templates were developed (these are available upon request to the authors) for the following five high-volume analytes (immunological methods) selected for the study: troponin I, troponin T, glycated hemoglobin (HbA_{1c}), prostate-specific antigen (PSA) and D-dimer. All of these analytes have clinical decision limits, and the members of our WG had experienced large lot-to-lot variations for some of the analytes.

Literature review

To identify publications relevant to this project, a systematic literature search of PubMed was performed in January 2019 and repeated in February 2021. The search terms used were "diagnostic reagent kits", "lot", and "lot-to-lot variation", which identified 376 papers. Title and Abstract screening was performed after removing one duplicate, which left 16 papers remaining from the originally retrieved publications. Twelve papers remained after full-text reviews performed by two of the authors [1–3, 6, 7, 10–15]. One guideline and one standard were additionally identified by the authors [5, 16]. Further information about the literature search is provided in Supplementary Table 1.

Determination of APS of lot-to-lot variations for the selected analytes

Sixty Norwegian medical laboratories were asked which APS they would use for assessing reagent lot-to-lot variations for each of the selected analytes (HbA_{1c}, D-dimer, PSA, troponin T and troponin I) at different levels by including the following question in a questionnaire:

> Assuming the reagent lot in use allows accurate analyte measurements in patients, what do you think the acceptance limit should be for the various analytes at two predefined levels; that is, what is the maximum bias (as a percentage) between the reagent lot in use and the new reagent lot that would make the new lot acceptable to you (without significant adjustments such as factorization or changing the reference intervals)?

The practical procedure proposed for evaluating lot-to-lot variations is described in Table 1, and the rationale behind some of the procedures are explained in more detail below.

The chosen APS are pragmatically based on a combination of state-of-the-art and expert opinions by the authors, where state-of-the-art means the performance of MPs already on the market and in use in routine laboratories. All recommended APS are listed in Table 2.

Patient samples and analyte concentrations

Patient samples collected within the laboratory and received for analysis should be used for testing new reagent lots, since internal analytical quality control materials often have a different matrix and therefore may lead to incorrect conclusions when evaluating reagent lot changes [2, 14, 15, 17]. The analyte concentrations should be close to the clinical decision limits and/or close to the upper or lower reference limits as indicated in Table 2. Some individual samples can exhibit behaviors that differ from the majority of patient samples, such as due to specific effects [5]. If this is suspected, more patient samples should be examined. A laboratory should use the templates to provide

Table 1: The proposed main steps for each laboratory in the lot-to-lot evaluation process.

1. Select the *a-priori* APS (see Table 2).
2. Calibrate the new reagent lot as per specifications.
3. Select patient samples in the specified concentration intervals (see Table 2). Avoid patient samples that are haemolyzed or have an increased concentration of bilirubin or lipemia (turbid samples). Check that there is enough sample material to analyze the recommended number of replicates. If one patient sample does not provide enough material for the recommended number of replicates, multiple patient samples can be pooled.
4. Analyze the patient samples using the existing reagent lot and the new reagent lot. Perform these analyses as closely as practicable in time and avoid unnecessary storage of the samples.
5. Insert the required data and the measured concentrations for all proposed levels in the spreadsheet prepared for the lot-to-lot evaluation for that specific analyte. The mean percentage difference between the patient samples analyzed using the new reagent lot and the reagent lot number in use along with the corresponding 90% CI will then be calculated.
6. Assess the results. If the mean percentage difference between the existing reagent lot and the new reagent lot including the 90% CI for all samples and concentration intervals is within the acceptance limit, the new reagent lot can be approved and used. If the mean percentage difference is within the acceptance limit but the 90% CI crosses it, the laboratory should analyze more samples and then combine the results. In addition, the cumulative changes should be monitored and included in the assessments of the results.
7. If the mean percentage difference between the existing reagent lot and the new reagent lot including the 90% CI is outside the acceptance limit, the new reagent should not be approved and used, and preagreed actions should be taken.

APS, analytical performance specifications; CI, confidence interval.

Table 2: Recommended concentrations, minimum number of replicates and acceptance criteria for the five selected analytes when testing new reagent lots.

Analyte	Concentration interval of the patient samples to be examined	Minimum number of replicates	Acceptance criterion for the mean difference between two lots, %
hs Troponin I (sample 1) (Abbott)	2–5 ng/L	6	20
hs Troponin I (sample 2) (Abbott)	20–32 ng/L	6	10
hs Troponin I (sample 3) (Abbott)	150–250 ng/L	6	10
hs Troponin I (sample 1) (Siemens Healthineers)	2–5 ng/L	6	20
hs Troponin I (sample 2) (Siemens Healthineers)	35–58 ng/L	6	10
hs Troponin I (sample 3) (Siemens Healthineers)	150–250 ng/L	6	10
hs Troponin T (sample 1) (Roche Diagnostics)	5–10 ng/L	6	20
hs Troponin T (sample 2) (Roche Diagnostics)	11–18 ng/L	6	10
hs Troponin T (sample 3) (Roche Diagnostics)	150–250 ng/L	6	10
HbA_{1c} (sample 1)	40–50 mmol/mol	6	5
HbA_{1c} (sample 2)	70–80 mmol/mol	6	5
PSA (sample 1)	0.1–0.5 µg/L	6	30
PSA (sample 2)	3.0–5.0 µg/L	6	20
PSA (sample 3)	10.0–20.0 µg/L	6	20
D-dimer (patient pool)	0.4–1.0 mg/L FEU	12	20

HbA_{1c}, glycated haemoglobin; PSA, prostate-specific antigen.

data about the laboratory, instrument and MP, lot number for the calibrator, the reagent lot number in use and the new reagent lot number, all results measured using the reagent lot in use and the new reagent lot, and their own evaluations of their results. For each sample, the mean difference (as a percentage, along with the 90% confidence interval [CI]) between the concentration in the sample analyzed with the reagent in use and the new reagent is calculated and displayed in a figure [18]. The templates also include a figure showing the cumulative percentage difference between multiple lot changes. In the present study, three different concentrations were chosen for evaluating lot changes for the troponins and PSA, while two were chosen for HbA_{1c} and one for D-dimer (Table 2).

Number of replicates of the patient samples

Sufficient statistical power can be attained either by analyzing a relatively large number of samples or a smaller number of samples in multiple replicates. In Norway many laboratories use the second approach because it can be difficult to find enough samples at the required levels. Commonly, two or three samples are selected and analyzed in six replicates with the two reagent lots. The practice is probably derived from a study from 2005 using a related protocol [19]. The number of replicates in this project was determined using power-function curves for equivalence testing [20]. The chosen power-function curves can be used to identify if the 90% CI for

the difference between the two means exceeds the acceptance limit. The power-function curves also provide information about the probability of false rejections, which should be acceptably low. The chosen APS were transformed into the number of standard deviations for the MS. The standard deviations for the MSs were based on the repeatability for the internal analytical quality control obtained for the analytes at some of the university hospital laboratories represented in the WG. If the repeatability varied for different MSs, the highest value was chosen. Finally, the number of replicates was pragmatically chosen by the members of the WG based on power-function curves [20] in the context of the chosen APS, repeatability of the MPs, probability of error detection and probability of false rejections. Six replicates were proposed for the troponins, HbA_{1c} and PSA, and 12 for D-dimer (Table 2). The mean percentage difference between two consecutive reagent lot numbers (including the 90% CI) should be within the acceptance criterion for the result to be evaluated as being within the APS.

Participants

In February 2020, all 60 Norwegian medical laboratories were invited to send their results from each lot change for each of the chosen analytes to Noklus by filling in the templates. Noklus would continuously evaluate the lot-to-lot variations of the MPs. Grouping of the MSs for each laboratory into MPs depends on the analyte in question (e.g., D-dimer is grouped by the assay type). The mean lot-to-lot variation for each MP is monitored and compared with other MPs. Participants receive a yearly report presenting anonymized results from all participants, and the first report was sent to laboratories in January 2021. When participants experience unexpected results, they can contact Noklus for information about the lot in question. As the number of results increases, the intention is to communicate the results to both the professional community and instrument manufacturers.

Results

Eighteen of the 60 laboratories responded to the questionnaire on APS for lot evaluation (Figure 1). There was a wide variation in the suggested APS. The responses to the questionnaire (Figure 1) and consensus within the WG were used to establish APS at important decision limits (Table 2).

Evaluating reagent lot changes in the individual laboratories

As at September 2021, 27 of the 60 Norwegian laboratories were participating in the national surveillance program; examples of the results obtained are presented below.

In the templates, figures for each lot change are automatically generated to be evaluated by each laboratory, which is exemplified for troponin I in Figure 2A. The mean, standard deviation, coefficient of variation and standard error of the mean are calculated for each sample and each concentration, and the mean difference and the corresponding 90% CI is shown in a diagram. Each laboratory also evaluates the cumulative differences between several consecutive lot changes on the same instrument each time they enter results into the Excel template, as exemplified for troponin I in Figure 2B, and the cumulative difference is calculated for each instrument, analyte and concentration by adding successively the effect and direction of lot changes. The four different lot changes shown in Figure 2A include a new reagent lot that could be accepted according to the proposed APS (5 September 2020), inconclusive

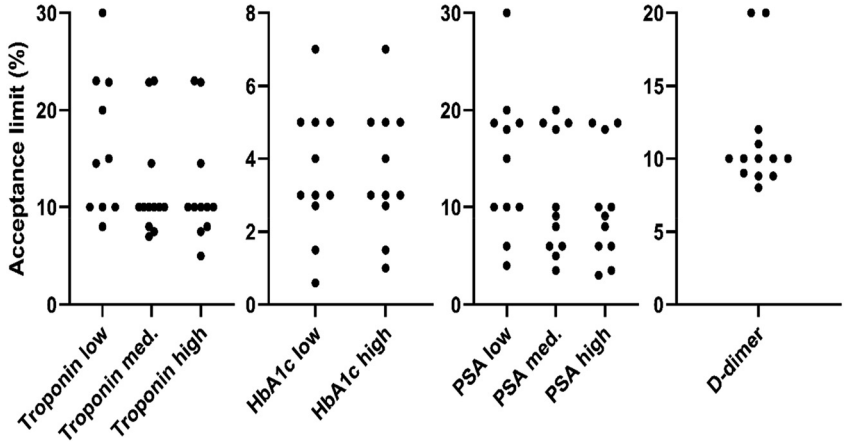

Figure 1: Lot-to-lot acceptance limits (systematic deviations between two lots) suggested by the 18 Norwegian laboratories.
The limits for troponin T and I for Roche Diagnostics, Abbott and Siemens Healthineers instruments are merged. The low, medium (med.) and high concentrations were 3, 15 and 100 ng/L, respectively, for troponin T with Roche Diagnostics instruments; 3, 50 and 300 ng/L for troponin I with Siemens Healthineers instruments; 3, 30 and 200 ng/L for troponin I with Abbott instruments; and 0.1, 4 and 10 µg/L for prostate-specific antigen (PSA). The low and high concentrations were 48 and 75 mmol/mol for HbA_{1c}, and the concentration interval was 0.5–1.0 mg/L for D-dimer.

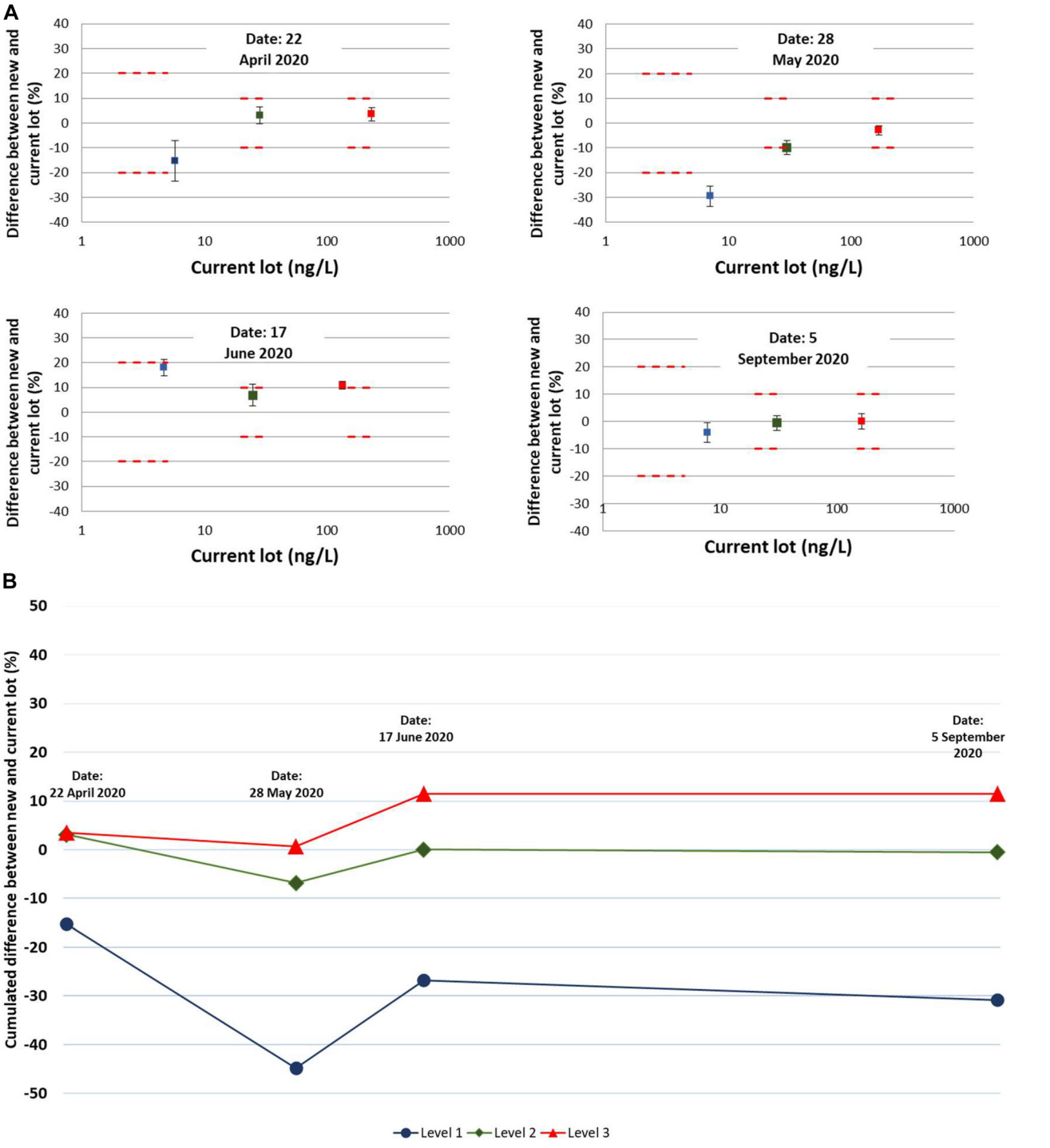

Figure 2: Results from the evaluation of four troponin I reagent lot changes using an Abbott instrument.
(A) The x-axis is the concentration of troponin I as measured using the current reagent lot, and the y-axis is the percentage difference between the new reagent lot and the current reagent lot in use. The squares with the vertical lines represent the mean difference with the 90% confidence interval (CI) for three different troponin I concentration intervals (level 1: 2–5 ng/L, level 2: 20–32 ng/L, level 3: 150–250 ng/L). The dashed lines are the acceptance limits for troponin I at the different levels. (B) Cumulative percentage difference on the y-axis for troponin I for the same four lot changes, including the dates of analyses.

results where the 90% CI includes the APS (22 April and 17 June 2020) and one lot change where the difference between the two reagent lots exceeds the APS (28 May 2020). As seen for the low concentration in Figure 2A, a sample can be used in the lot evaluation even if the laboratory is not able to collect a sample in the predefined interval.

Results from the national surveillance program for monitoring reagent lot verifications

Individual results from all laboratories were grouped according to MPs and concentration intervals. For each MP the absolute mean value and the maximum and minimum difference between two consecutive reagent lots are calculated. When results are presented, also the number of results included in the absolute mean difference is presented. Observed lot-to-lot variations were compared within each MP, as exemplified in Figure 3 (lot number codes are listed in Supplementary Table 2) for troponin T (hs troponin T, Roche Diagnostics, Mannheim, Germany) performed using different instruments in the Cobas family. Figure 3 shows the results from an evaluation of the same reagent lot change on two different instruments (from reagent lot A to reagent lot B). There was no clinically significant lot-to-lot variation for one instrument, while the results were inconclusive for the other instrument (i.e., the 90% CI includes the acceptance limit).

Results from different methods can also be compared, as exemplified in Figure 4 (lot number codes are listed in Supplementary Table 3) for D-dimer at 0.4–1.0 mg/L FEU for four different MPs (Hemosil D-dimer HS, Instrumentation Laboratory, Bedford, MA, USA; Innovance D-dimer, Siemens Healthineers, Erlangen, Germany; STA-Liatest D-dimer, Diagnostica Stago, Paris, France; and Cobas Tina-quant D-dimer, Roche Diagnostics) and in Figure 5 (lot number codes are listed in Supplementary Table 4) for PSA at 3–5 µg/L for four different MPs (Alinity, Abbott, Chicago, IL, USA; Architect, Abbott; Centaur, Siemens Healthineers; and Cobas, Roche Diagnostics).

The suggested APS were achievable for most of the analytes and MPs included in the program, yet for several of the analytes there were individual results exceeding the APS (Figures 3–5). The preliminary results also indicate that the lot-to-lot variations differed between different MPs (Figures 4 and 5). All of the results received by the program up to May 2021 are summarized in Table 3.

Discussion

This study analyzed a simple procedure proposed for the national surveillance of reagent lot evaluations. The procedure includes evaluating the effect of cumulative lot changes. The APS for the selected analytes (troponin I, troponin T, HbA_{1c}, PSA and D-dimer) were first established, and then practical procedures for testing lot-to-lot variations in individual laboratories were implemented. Some of our preliminary results have been presented here to illustrate the plan for using the model in a national surveillance program.

The CLSI EP26-A guideline recommends that APS be set within each laboratory [5]. Based on the results from the survey presented in Figure 1, where laboratories were asked for APS for the selected analytes, and also on our

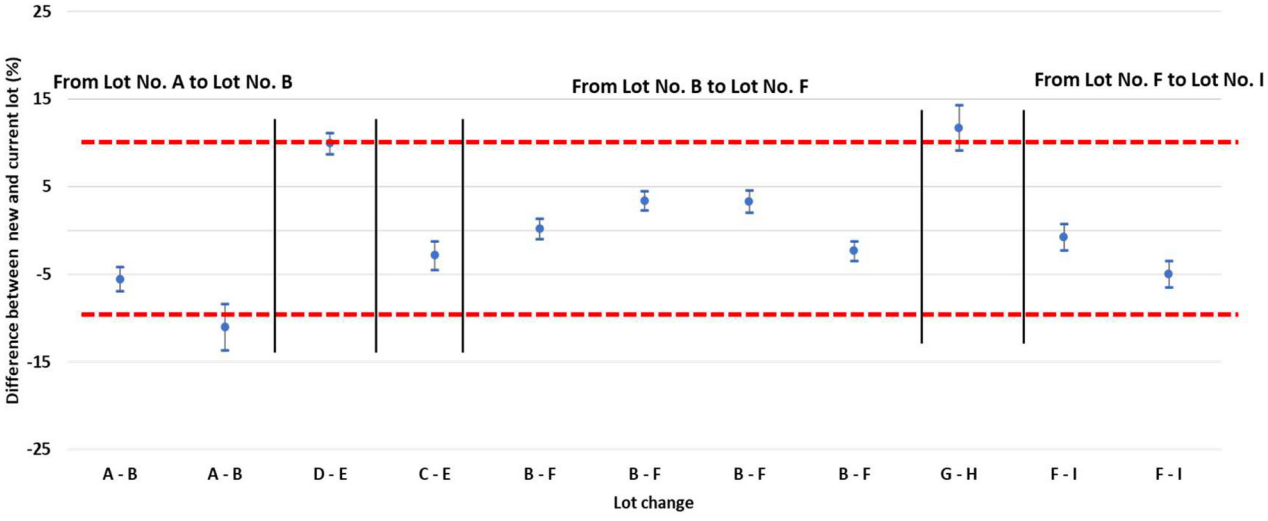

Figure 3: Results for 11 troponin T reagent lot changes analyzed in seven different laboratories using eight Cobas instruments (Roche Diagnostics) over the concentration interval of 11–18 ng/L.
The y-axis is the mean percentage difference between two consecutive reagent lots, and the x-axis is the different lot numbers tested (coded from A to I). The circles and bars represent the mean percentage differences between pairs of reagent lots and the corresponding 90% CIs. The dashed lines are the acceptance limits (±10%). Lot number codes are listed in Supplementary Table 2.

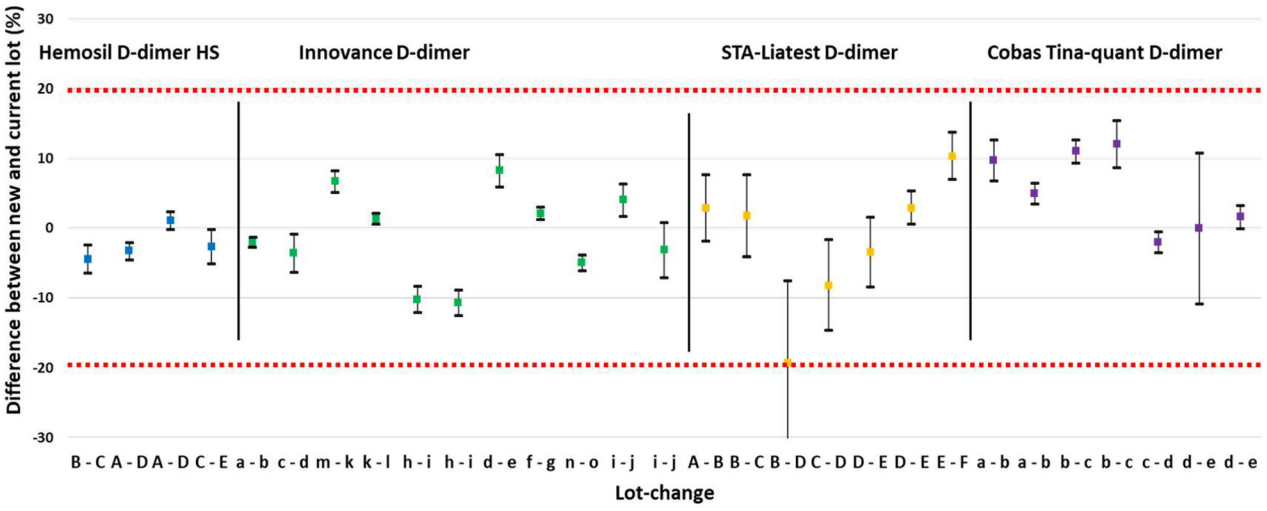

Figure 4: Results for 29 D-dimer reagent lot changes analyzed in 13 different laboratories using 15 different instruments with four different measurement procedures (MPs) over the concentration interval of 0.4–1.0 mg/L FEU.
The letters on the x-axis represent the different lot numbers for the different MPs, and the y-axis is the percentage difference between two reagent lots. The squares and bars represent the mean percentage differences between pairs of reagent lots and the corresponding 90% CIs. The dashed lines are the acceptance limits (±20%). Lot number codes are listed in Supplementary Table 3.

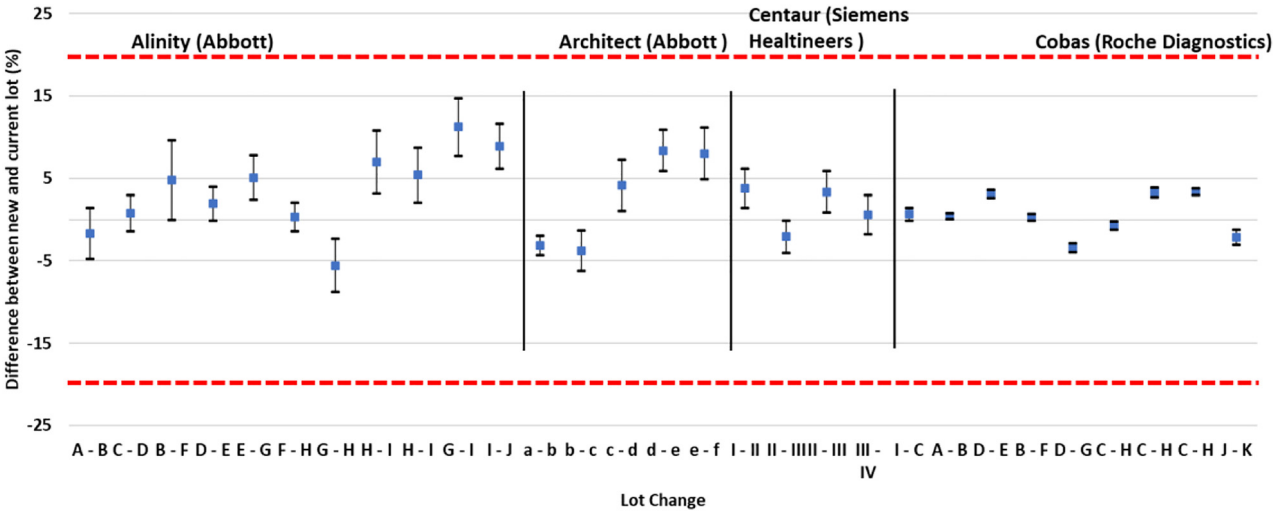

Figure 5: Results for 29 prostate-specific antigen (PSA) reagent lot changes analyzed in 14 different laboratories using 15 different instruments with four different measurement procedures (MPs) over the concentration interval of 3.0–5.0 µg/L.
The letters on the x-axis represent the different lot numbers for the different MPs, and the y-axis is the percentage difference between two reagent lots. The squares and bars represent the mean percentage differences between pairs of reagent lots and the corresponding 90% CIs. The dashed lines are the acceptance limits (±20%). Lot number codes are listed in Supplementary Table 4.

experience from the workshop for medical laboratories, it seems to be difficult for laboratories to set these specifications. This suggests that it would be better if common APS for each analyte within a region or country are established by expert groups [21]. Harmonized APS would also make it easier to compare results from different laboratories. Some of the APS might be considered as broad. When we have more data, the APS will be evaluated taking performance specifications based on biological variation differentiated into bias and total error into account, and it will be possible to discuss the clinical implications of the chosen APS for lot-to-lot variation and if needed, revise the recommended APS.

There have been different proposals on how to test for lot-to-lot variations [5, 6, 10, 11, 15], with procedural differences such as in how the APS are set, sample selection

Table 3: A summary of all the results received by the program up to May 2021.

Analyte (no. of lot changes)	Concentration interval (unit) Mean difference (minimum–maximum), %		
Troponin I (Abbott) (29[a])	2–5 ng/L	20–32 ng/L	150–250 ng/L
	11.9 (1.8–33.2)	4.5 (0.0–12.1)	3.2 (0.0–10.9)
Troponin T (Cobas, Roche Diagnostics) (11)	5–10 ng/L	10.7–17.5 ng/L	150–250 ng/L
	7.6 (1.4–17.7)	5.1 (0.2–11.7)	2.6 (0.3–6.4)
HbA_{1c}[b] (28)	40–50 mmol/mol	70–80 mmol/mol	
	2.2 (0.0–6.0)	1.8 (0.0–6.9)	
PSA (29)	0.1–0.5 µg/L	3–5 µg/L	10–20 µg/L
	3.6 (0.0–11.1)	3.7 (0.3–11.2)	4.3 (0.1–11.6)
D-dimer (29)	0.4–1.0 mg/L FEU		
	5.5 (0.0–19.3)		

Mean absolute percentage difference between two reagent lots, the minimum absolute percentage difference, the maximum absolute percentage difference and the number of lot changes. The results are for different laboratories, instruments and MPs. [a]Only 25 laboratories verified new reagent lots for the lower concentration interval. [b]Results from HPLC-methods represents new columns. HbA_{1c}, glycated haemoglobin; PSA, prostate-specific antigen.

and number of samples, whether or not extreme results should be included, and the statistics utilized in comparisons. Challenges associated with implementing these procedures in laboratories with marked differences in size and competence may critically limit the utility of these protocols. To be able to compare results from different laboratories and include them in a national surveillance program, it is important to harmonize the lot-to-lot evaluation procedure. As a start, we have chosen analytes, reagents and MPs for which testing for lot-to-lot variation is especially important because of the clinical decision limits used for diagnosis and monitoring. The procedure proposed by our WG for testing lot-to-lot variations in individual laboratories represents only one of several ways that a laboratory can meet the requirements in ISO 15189 when determining whether a new reagent lot meets the APS prior to use [16].

It might be problematic that our proposed procedure used only three patient samples (for D-dimer one sample) to represent all patient samples. The CLSI EP26-A guideline recommends always including at least three patient samples, which may be dispersed among the target concentrations being evaluated [5]. However, evaluating multiple concentrations is proposed since it cannot be ruled out that lot-to-lot variations are influenced by the concentration, and this approach also makes it less likely that an unusual sample will result in an erroneous conclusion. For D-dimer we proposed only a single concentration interval, because the clinical decision limit for D-dimer is the most important one used in algorithms for excluding venous thromboembolism.

We recommend that laboratories analyze additional samples when the results are inconclusive; that is, when the 90% CI crosses the acceptance limit. The template includes the possibility to add more samples while still retaining the information in the plot of cumulative differences.

Our proposed procedure for lot testing represents a pragmatic compromise between having sufficient power to detect important lot changes while still ensuring that the procedure is feasible to perform in all laboratories. As can be seen in some of the examples, there are cases were individual laboratories get diverging results when evaluating the same lot change. This could be due to differences in e.g., local calibrations. Even if some of the results are exceeding the recommended APS, to our knowledge none of the laboratories have rejected a new reagent lot. Also, we have not systematically evaluated whether the lot changes observed in this study correspond to actual changes for a larger number of patient samples.

If the results from an evaluation of a new reagent lot exceed the APS, the laboratory must try to determine the underlying reason. If it is concluded that there is an important bias between the new reagent lot and the reagent lot in use, the manufacturer should be asked to supply an alternative reagent lot. The action to take if an alternative lot is not available is left to the laboratories, because the optimal solution might differ for various settings and analytes. The options depend on various factors, such as the analyte involved, how far the results are outside the APS and the results obtained when analyzing previous lots. To supplement the testing of a new reagent lot and to follow its stability over time, for many analytes it is useful to also monitor the moving patient median or average value [3, 22]; however, such information will only be available after the new lot has been used for a sufficient length of time.

Results presented here underline the importance of the procedure used to evaluate new reagent lots being easy to perform and therefore feasible in both small and large medical laboratories. Noklus has also found it necessary to offer support and to answer practical questions from the participants in a timely manner. So far, the experience is that the method information received by Noklus from the laboratories is often incomplete and that additional information needs to be requested. Creating a uniform way of reporting lot numbers specific to different manufacturers would improve the reliability of the information about reported lot numbers.

The willingness of laboratories to share results makes it possible to perform an overall evaluation of lot-to-lot variations. All of our results were obtained from routine laboratories rather than from optimally designed studies. The aggregated results as presented in Table 3 give an impression of the expected lot-to-lot variation for an analyte, including the maximum variation found in routine laboratories between reagent lots. As we accumulate more data (Figures 4 and 5), laboratories could find this information useful for deciding about which instruments to purchase or which MPs to use and might also give manufacturers a stronger incentive to work on reducing lot-to-lot variations. It is challenging to make routine laboratories agree on the same procedure for lot-to-lot evaluations, but the advantages of performing a centralized follow-up and compiling data on several analytes and MPs outweigh the drawbacks. The individual testing of one new reagent lot according to our model was made simple by adopting a pragmatic approach in which the model fitted as many routine laboratories as possible. The results compiled in a national surveillance program provide important knowledge about the lot-to-lot variations in medical laboratories performing routine procedures. The model will be evaluated based on the results from the participating laboratories, and then modified if needed. The experiences during this study will be valuable when adding other analytes to the program.

In conclusion, this study found that sharing information from lot evaluations may allow simplification of lot testing in individual laboratories and has provided real-world information about lot-to-lot variations for different MPs.

Research funding: None declared.

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

Competing interests: The authors have no conflict of interest to declare.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References

1. Thaler MA, Iakoubov R, Bietenbeck A, Luppa PB. Clinically relevant lot-to-lot reagent difference in a commercial immunoturbidimetric assay for glycated hemoglobin A1c. Clin Biochem 2015;48:1167–70.
2. Kristensen GB, Christensen NG, Thue G, Sandberg S. Between-lot variation in external quality assessment of glucose: clinical importance and effect on participant performance evaluation. Clin Chem 2005;51:1632–6.
3. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SK. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem 2013;59:1187–94.
4. Haagensen K, Collinson P, Asberg A, Aakre KM. How does the analytical quality of the high-sensitivity cardiac troponin T assay affect the ESC rule out algorithm for NSTEMI? Clin Chem 2019;65:494–6.
5. CLSI. User evaluation of between-reagent lot variation; Approved Guideline. CLSI document EP26-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2013.
6. Katzman BM, Ness KM, Algeciras-Schimnich A. Evaluation of the CLSI EP26-A protocol for detection of reagent lot-to-lot differences. Clin Biochem 2017;50:768–71.
7. Kim HS, Kang HJ, Whang DH, Lee SG, Park MJ, Park JY, et al. Analysis of reagent lot-to-lot comparability tests in five immunoassay items. Ann Clin Lab Sci 2012;42:165–73.
8. Holzel W. Analytical variation in immunoassays and its importance for medical decision making. Scand J Clin Lab Invest Suppl 1991;205:113–9.
9. Algeciras-Schimnich A. Tackling reagent lot-to-lot verification in the clinical laboratory AACC. AACC; 2014. Available from: https://www.aacc.org/publications/cln/articles/2014/july/bench-matters.
10. Don-Wauchope AC. Lot change for reagents and calibrators. Clin Biochem 2016;49:1211–2.
11. Hofmans M, Oyaert M, De Schrijver P, Nobels F, Van Hoovels L. Clinical laboratories have a critical role in test strip lot management in glucose point-of-care testing. Clin Chem Lab Med 2016;54:e155–9.
12. Liu J, Tan CH, Loh TP, Badrick T. Detecting long-term drift in reagent lots. Clin Chem 2015;61:1292–8.
13. Martindale RA, Cembrowski GS, Journalt LT, Crawford JL, Tran C, Hofer TL, et al. Validating new reagents: roadmaps through the wilderness. Lab Med 2006;37:347–51.
14. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. Clin Chem 2011;57:76–83.
15. Thompson S, Chesher D. Lot-to-lot variation. Clin Biochem Rev 2018;39:51–60.
16. ISO. Medical laboratories — requirements for quality and competence. ISO 15189. Geneva, Switzerland: International Organization for Standardization; 2012.
17. Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem 2006;52:553–4.

18. Altmann DG. Practical statistics for medical research. London, England: Chapman & Hall/CRC; 1991.
19. Kallner A, Khorovskaya L, Pettersson T. A method to estimate the uncertainty of measurements in a conglomerate of instruments/ laboratories. Scand J Clin Lab Invest 2005;65:551–8.
20. Asberg A, Solem KB, Mikkelsen G. Determining sample size when assessing mean equivalence. Scand J Clin Lab Invest 2014;74: 713–5.
21. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53: 833–5.
22. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, van Rossum HH, Loh TP, et al. Patient-based real-time quality control: review and recommendations. Clin Chem 2019;65: 962–71.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2021-1262).