External Quality Assessment of Point-of-Care Methods: Model For Combined Assessment of Method Bias and Single-Participant Performance by the Use of Native Patient Samples and Noncommutable Control Materials

Anne Stavelin,1,2 Per Hyltoft Petersen,1 Una Ø. Sølvik,2 and Sverre Sandberg2,3

BACKGROUND: An important objective in external quality assessment (EQA) is to evaluate systematic deviations between methods. However, this is not possible when noncommutable control materials are used. The aim of this study was to develop an EQA model that incorporates a method bias evaluation using native patient samples EQA schemes in which noncommutable materials are used.

METHODS: The model was applied twice in a point-of-care (POC) international normalized ratio survey among 1341 and 1578 participants. To estimate bias, about 100 native patient samples for each POC method were analyzed by a selected group of “expert” primary healthcare centers and on a designated comparison method. In addition, the expert centers as well as all the other EQA participants analyzed 2 noncommutable control materials, and method-specific target values were established. Both method bias and the deviation of a single-participant result from the method target value were evaluated against analytical quality specifications, making combined assessment possible. The best-case scenario occurred when both results were within the quality specifications.

RESULTS: Two POC methods fulfilled the quality specifications for bias, whereas one did not. The best-case scenario was achieved by more than 90% of the participants using the methods with no bias, whereas none of the participants using the method with unacceptable bias achieved this result.

CONCLUSIONS: We propose an EQA model for which the bias of POC methods can be evaluated in situations in which commutable materials are not available.

An optimal external quality assessment (EQA)4 scheme is characterized by the use of commutable, stable, and homogeneous control materials with the same matrix as patient samples (1, 2). The target value should be established by a reference method and the control samples should be handled similarly to patient samples. In many cases, however, optimal conditions are difficult to achieve. In a recent review different EQA schemes were classified into 6 categories (2), in which category 1 was characterized by commutable samples with reference target values and categories 5 and 6 were characterized by noncommutable samples with peer-group target values. In these schemes, different aspects of the analytical quality of both the methods and each individual participant can be assessed (2, 3). Efforts have been initiated to improve the harmonization between laboratories when there are no reference measurement procedures available (4). A presumption for performing EQA with noncommutable samples with method-specific peer-group target values is that the participant’s bias obtained when a noncommutable sample is used should reflect the participant’s bias obtained when a commutable material is used.

With the expanding use of point-of-care (POC) testing in decentralized locations, such as surgical departments, general practitioner offices, nursing homes, and pharmacies (5), there is a need to establish EQA in these environments. Several guidelines (6–9) recommend that POC users should participate in an EQA scheme to improve measurement quality. This process, however, is a challenge for the EQA organizers owing to the large numbers of control samples needed, the possible requirement of different control materials for different POC instruments, and the lack of tradition for QC systems in these locations. Nevertheless, studies...
have shown that participation in EQA for POC testing is a useful tool for improvement of performance (10–13).

Good measurement quality is essential in determinations of international normalized ratio (INR) for safe and efficient anticoagulation treatment with vitamin K antagonists, and participation in an EQA scheme is recommended (9). However, a recent study (14) has shown that in most European countries EQA for POC INR testing is not provided, primarily because of the lack of a suitable control material. The study also showed that different types of EQA schemes are provided and all have advantages and limitations. The main drawback of the most common schemes is that they use noncommutable control samples (lyophilized materials), and alternative approaches have therefore been developed (14).

To assess a possible systematic deviation between methods, it is common to conduct method-comparison studies (15). These studies are, however, not performed on a regular basis and are often large and time-consuming. Such evaluations are, for example, performed by SKUP (the Scandinavian evaluation of laboratory equipment for primary healthcare) (16), where POC instruments are evaluated once before entering the Scandinavian market. However, an evaluation performed on a regular basis is needed to monitor the analytical quality of these instruments over time, including lot-to-lot reagent variation.

The aim of the present study was to develop an EQA model that combines an assessment of POC method bias with single-participant performance. The model should be easy to carry out on a regular basis and is particularly aimed for constituents for which the EQA control material is noncommutable.

### Material and Methods

**PROPOSED MODEL**

The principle for the proposed model is shown in Fig. 1 and is outlined below:

1. **Estimation of method bias by the use of native patient samples.** Some expert primary healthcare centers (e.g., 20 for each POC method) are selected and a comparison method is chosen. The designated comparison method should ideally be a reference method or alternatively be a well-established laboratory method which is traceable to a reference measurement procedure. A few (e.g., 5) native patient samples are analyzed at each primary care center and with the designated comparison method. The mean of all differences (about 100) between the POC method and the comparison method is calculated to give an estimate of the POC method bias. The native patient samples should span a reasonable range. If the designated comparison method is not a reference method, the results could be adjusted with results from certified reference materials. The POC reagent lot number should be registered and if lot-to-lot variation is present, this must be taken into consideration when interpreting the results.

![Fig. 1. Principle of the model.](image_url)
2. Establishment of method-specific target values by the use of noncommutable control materials. In the same time period, noncommutable control materials are distributed to all EQA participants and to the expert centers. A method-specific target value for each POC method is calculated on the basis of results from the expert centers. Each single-participant result can then be compared with this target value. As a result, there is a traceability chain from the comparison method to the POC method through the expert centers and thereby to each individual EQA participant.

3. Combined assessment of method bias and single-participant performance. The deviation between one single-participant result and the method-specific target value is calculated and plotted against the estimated method bias in the same figure. (Fig. 2). Separate analytical quality specifications for the POC method bias and the single-participant result should be used, and a combined assessment with different interpretations is possible (Fig. 2, fields A to I).

APPLICATION OF THE MODEL ON POC INR

Selection of expert primary healthcare centers. The selection criteria for the expert primary healthcare centers were that they had to use one of the most common POC INR methods in Norway and have good analytical quality and skilled personnel with POC INR testing experience. The most commonly used POC instruments were the CoaguChek XS Plus® (Roche Diagnostics), Simple Simon® (Zafena AB), and Thrombotrack® (Axis-Shield). Good analytical quality was defined as results within the acceptability limits for “good performance” (14) for both controls 1 and 2 in the last 2 EQA schemes prior this study. The selection of centers with skilled personnel was performed by the laboratory advisors in the Norwegian Quality Improvement of Primary Care Laboratories (NOKLUS). A total of 128 primary healthcare centers were invited to participate in this study, and 72 agreed to participate in the first survey in May to June 2010. Of these 72 centers, 69 agreed to also participate in the second survey in May to June 2011. The aim was to get approximately 100 comparisons between each POC method and the comparison method. The number of samples and participating centers for both surveys are shown in Table 1.

COLLECTION AND MEASUREMENTS OF NATIVE PATIENT SAMPLES

Each primary care center selected 4 or 5 patients on long-term anticoagulation treatment scheduled to be seen for a follow-up visit with their general practitioner within the 2 given weeks in each survey. Each patient received written and oral information about the study and gave oral informed consent to participate. The Regional Committee for Medical and Health Research Ethics in Norway approved the study. The centers analyzed the native patient samples by their own POC method according to their routine procedures, and the reagent lot number was registered. Capillary whole blood was used for CoaguChek XS Plus, whereas citrated whole blood (3.2% or 3.8% sodium citrate) was used for Simple Simon and Thrombotrack. An extra tube with citrated whole blood was obtained from each of the patients and sent by ordinary mail to the hospital laboratory for comparison. The samples were analyzed by the designated comparison method the day of arrival.

DESIGNATED COMPARISON METHOD

The most commonly used INR method in Norwegian hospitals was chosen to be the designated comparison...
method. The method used at the hospital laboratory (Haraldsplass Deaconess Hospital, Bergen, Norway) was the STA R Evolution® (Diagnostica Stago) and the SPA reagent (Diagnostica Stago), which is a combined rabbit brain thromboplastin with adsorbed bovine plasma (Owren method) \((17)\). The final sample dilution is 1:21. One batch of reagent was used for each survey, and the international sensitivity index values were 0.89 and 0.91, respectively. The calibrators were traceable to WHO International Reference Preparation RBT/90 \((18)\). The comparison method was calibrated at the start of each survey. The day-to-day reproducibility CV of the method was 1.4% (level 1.0 INR) and 2.0% (level 3.0 INR). Three fresh-frozen control plasmas in 3 INR levels (used in EQA for hospitals) with assigned target values from the median of Norwegian laboratories using the same method \((n = 55)\) were analyzed at the start, during, and at the end of each survey. This was done to ensure that the designated comparison method represented “a typical Norwegian hospital method,” and all INR results from the comparison method were adjusted according to these control plasma results. The results from the designated comparison method were adjusted by the regression equations: 

\[
y = 1.024x + 0.013 \quad \text{and} \quad y = 1.011x + 0.027 \]

for the 2010 and the 2011 survey, respectively.

**Table 1. Bias for the different POC INR methods in the 2010 and 2011 surveys.**

<table>
<thead>
<tr>
<th>Method</th>
<th>CoaguChek XS Plus</th>
<th>Simple Simon</th>
<th>Thrombotrack</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of expert primary care centers</td>
<td>23</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>No. of included patient samples</td>
<td>99</td>
<td>105</td>
<td>113</td>
</tr>
<tr>
<td>INR range of patient samples</td>
<td>1.25 to 4.15</td>
<td>1.28 to 4.01</td>
<td>1.23 to 6.11</td>
</tr>
<tr>
<td>Method bias, %</td>
<td>-0.4</td>
<td>-1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Method bias, 95% CI</td>
<td>-2.1 to 1.2</td>
<td>-2.6 to 0.2</td>
<td>-0.5 to 2.7</td>
</tr>
</tbody>
</table>

**ESTIMATION OF POC INR METHOD BIAS**
The mean of all differences (in percentage) between the results from the POC INR method and the designated comparison method was calculated to give an estimate of the method bias. Totals of 44 and 29 samples were excluded in the 2010 and 2011 surveys, respectively. Reasons for exclusion were: samples older than 48 h, hemoglobin above 18 g/L for Thrombotrack samples, problematic sample collection, patients in the initial phase of anticoagulation treatment, and statistical outliers. The samples were analyzed for outliers according to Burnett \((19)\), and 1 outlier was excluded in 2011. The 48-h stability limit was the main reason for exclusion \((39 \text{ in } 2010 \text{ and } 25 \text{ in } 2011)\). INR sample stability is dependent on the instrument–reagent combination used, and each laboratory should therefore confirm its own system \((20\text{–}22)\). The 48-h stability limit was confirmed by 2 pilot studies before this study (data not shown).

**ESTABLISHMENT OF INR METHOD-SPECIFIC TARGET VALUES**
In the same time period during which the native patient samples were analyzed, noncommutable control materials were distributed to the expert primary healthcare centers and all the other participants in the EQA scheme for POC INR provided by NOKLUS. The total numbers of participants in the 2010 and 2011 surveys were 1341 and 1578, respectively. Different types of commercial lyophilized control plasma samples were distributed to the different types of POC INR instruments, TriniCHECK (Trinity Biotech) and OKP (MediRox AB) for Simple Simon, TriniCHECK and Control Plasma AK (Axis-Shield) for Thrombotrack, and UKNEQAS (the United Kingdom National External Quality Assessment Scheme) for Blood Coagulation provided samples for the CoaguChek XS Plus. The method-specific target values were calculated as the median INR values from the expert centers for each POC INR method, after exclusion of outliers (Table 2).

**COMBINED ASSESSMENT OF POC INR METHOD BIAS AND SINGLE-PARTICIPANT PERFORMANCE**
The deviation between each single-participant result and the INR method target value was plotted against the estimated method bias according to the proposed model. The Danish quality specification for a method bias of 6% \((23)\) was used. The quality specification for a deviating result from method target was set to 15%, which is the most commonly used quality specification for EQA for POC INR in Europe \((14)\). One of the likely reasons why this quality specification is wider than the specification for method bias is that it includes random error as well as the systematic error. The combined assessment was interpreted as described in Fig. 2. It is, however, important to notice that the quality specifica-
tions were set separately and that the evaluations were done separately, although they are presented in the same figure.

Results

ESTIMATED METHOD BIAS USING NATIVE PATIENT SAMPLES
The estimated bias for the different POC INR methods is shown in Table 1. The results were similar in 2010 and 2011. The CoaguChek and Simple Simon fulfilled the quality specification of bias of <6%, whereas the Thrombotrack exceeded this limit. Two and 4 different CoaguChek reagent lot numbers, and 3 and 2 different Thrombotrack reagent lots were used in the 2010 and the 2011 surveys, respectively, while only 1 Simple Simon reagent lot was used in each survey. There were no significant differences between lot numbers within 1 type of POC method. This result is illustrated for 5 CoaguChek lot numbers and 4 Thrombotrack lot numbers in Fig. 3A and C, respectively.

ESTABLISHMENT OF METHOD-SPECIFIC TARGET VALUES USING NONCOMMUTABLE CONTROL MATERIALS
Results for the 2010 and 2011 survey with 2 noncommutable control materials for the expert primary healthcare centers and for all participants using different types of POC INR methods are shown in Table 2. The median INRs from the expert centers (target values) were similar to the medians from all participants. The interlaboratory CV was similar or lower for the expert centers than for all participants. The portion of participants with results within the quality specification for a deviating result from the method-specific target value was between 97% and 98% for the CoaguChek and Thrombotrack users, and between 90% and 97% for the Simple Simon users for both control materials in the 2 surveys. In each survey several different reagent lot numbers were used by the CoaguChek and the Thrombotrack participants, whereas only 1 lot was used by the Simple Simon participants. There were significant differences between some of these lot numbers. The situation in which the native patient samples and the noncommutable control materials were analyzed on the same lot numbers is shown in Fig. 3. Significant differences were seen between the CoaguChek lots 3 and 4, between lots 4 and 5 for both controls in the 2011 survey (Fig. 3B), between Thrombotrack lots 6 and 7, and between lots 8 and 9 for both controls in both surveys (Fig. 3D).

COMBINED ASSESSMENT OF METHOD BIAS AND SINGLE-PARTICIPANT PERFORMANCE
An example of combined assessment is shown for 3 participants using different POC INR methods (Fig. 4). The best-case scenario was achieved by more than 90% of the CoaguChek and Simple Simon users, whereas none of the Thrombotrack users achieved this result. This is because the CoaguChek and Simple Simon ful-

Table 2. Results for the expert primary health care centers and for all participants using different types of POC INR methods, in the 2 surveys with 2 noncommutable control materials.

<table>
<thead>
<tr>
<th></th>
<th>Survey 2010</th>
<th></th>
<th>Survey 2011</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control 1</td>
<td>Control 2</td>
<td>Control 1</td>
<td>Control 2</td>
</tr>
<tr>
<td></td>
<td>Experts</td>
<td>All</td>
<td>Experts</td>
<td>All</td>
</tr>
<tr>
<td>CoaguChek XS Plus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of results</td>
<td>21</td>
<td>433</td>
<td>23</td>
<td>663</td>
</tr>
<tr>
<td>Median INR(^a)</td>
<td>3.0</td>
<td>3.0</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>CV, %(^b)</td>
<td>3.4</td>
<td>3.3</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Simple Simon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of results</td>
<td>23</td>
<td>204</td>
<td>20</td>
<td>190</td>
</tr>
<tr>
<td>Median INR(^a)</td>
<td>3.55</td>
<td>3.60</td>
<td>3.95</td>
<td>3.95</td>
</tr>
<tr>
<td>CV, %(^b)</td>
<td>4.2</td>
<td>7.1</td>
<td>4.2</td>
<td>6.2</td>
</tr>
<tr>
<td>Thrombotrack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of results</td>
<td>25</td>
<td>540</td>
<td>23</td>
<td>515</td>
</tr>
<tr>
<td>Median INR(^a)</td>
<td>2.30</td>
<td>2.30</td>
<td>2.50</td>
<td>2.55</td>
</tr>
<tr>
<td>CV, %(^b)</td>
<td>3.9</td>
<td>4.8</td>
<td>4.5</td>
<td>4.9</td>
</tr>
</tbody>
</table>

\(^a\) The median INR values established by the expert primary health care centers were used as the method-specific target values.

\(^b\) Interlaboratory CV.
filled the bias specification (meaning that the participants got results only in field A, B, and C), whereas the Thrombotrack did not fulfill the specification for bias (meaning that the participants got results only in field E, F, and I) (also see Supplemental Figs. 1–3 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue2).

Discussion

One of the goals for EQA programs is to include an evaluation of method bias (2). However, when noncommutable control samples are used, method bias evaluation is not possible. In this study, to assess both the methods and each single-participant performance, we used an EQA model that incorporated a method bias evaluation using native patient samples in an EQA scheme in which a noncommutable sample was used. The model was specifically aimed at POC methods but can also be used for other methods using peer group assessment. The model should be easy to carry out on a regular basis.

To our knowledge, there are no studies that describe the combination of method comparison with EQA schemes. However, Ross et al. (24) have developed a model which estimates the random and systematic matrix effects of an EQA control material compared to a native material. The target values, set by reference methods, on the control material can then be “corrected” for these matrix effects to assess the true-ness of methods. The main difference of this model from our model is that the “reference target values” in our model are set on the native material whereas in the model of Ross et al. the reference target values are set on the control material. Results of another study (3) suggested that a selected group of participants could estimate the method bias once a year by using fresh patient samples and then comparing the results with a reference method and communicating this information to all participants. Our model is an elaboration of this suggestion.

For the EQA organizers the proposed model might be simpler and easier to carry out than traditional method comparison studies in which all the measure-
ments (both by the POC methods and by the comparison method) are performed at the same location (15). EQA organizers will often have limited access to patients, and it can be difficult to perform method comparison studies in parallel with an EQA scheme using noncommutable control material. However, it is important to notice that the proposed model does not compete with traditional method evaluations studies, in which variables such as interferences, linearity, detection limits, and carryover are evaluated (3), but is a tool for estimating method bias on a regular basis.

One of the main challenges for all EQA organizers is to obtain commutable control materials (25). By implementing the present model, the EQA schemes become more useful without the need to distribute native samples to all participants. The model is therefore particularly suitable for POC methods, because there are often a large number of participants in these schemes. In addition, several different POC methods can be compared to the same designated comparison method, meaning that biases across the different methods can be compared, as shown for the INR methods in this study (Table 1).

Selection of expert primary healthcare centers is a key point in the proposed model. It is therefore important that the selection criteria are carefully chosen and that the EQA organizer has a system for selecting such centers. A presumption for the selection is that such centers perform the test correctly so that user errors are minimized. As a consequence, the interlaboratory variation should be lower or similar for all EQA participants. This goal was obtained for INR, as shown in Table 2. Another key factor in the estimation of method bias is the selection of a designated comparison method. The ideal would be to use a reference method (2), but for many constituents this would be expensive and practically difficult. Therefore, a well-established laboratory method traceable to a reference method is proposed as an alternative. For INR, it is known that different thromboplastins and methods give different INR values (26, 27). Therefore, the designated comparison method was adjusted so that it should have represented all laboratories in Norway with the same method. Thus, the POC INR methods were not compared only with one single laboratory method, but with “a typical Norwegian INR method.” A third important factor in the method bias estimation is the evaluation of results from different reagent lots. For the POC INR methods in this study there were no deviating reagent lots for any of the methods evaluated by the native samples, and no lots were excluded from the bias calculations.

It is known that noncommutable samples can give different lot-to-lot variation compared to commutable samples (28, 29). This was also the case for some lot numbers in our study, for which significant differences were seen for the noncommutable samples and not for the native patient samples (Fig. 3). Thus in the proposed model, it is possible both to monitor method
bias as well as lot-to-lot variation over time using native material (Table 1; Fig. 3A, C), and to evaluate whether the lot-to-lot variation using noncommutable control materials is similar to the lot-to-lot variation obtained with native material (Fig. 3). However, the model can detect only large differences between lot numbers using patient material, and for Thrombotrack (Fig. 3C, D) it cannot be excluded that if more native samples were analyzed, lot-to-lot differences also could have been detected in this material. This could have been evaluated further, especially in situations in which the results from the noncommutable material gave a suspicion of lot-to-lot variation. Information about substantial differences between lot numbers using native material should be given both to the manufacturer and to the participants. It is, however, difficult for the participant to deal with this situation, other than to change the lot. On the other hand such information is very useful for the manufacturer. If there is no significant difference between lots with the use of native material, but significant differences between lots with the noncommutable material, this is an indication that the control material is not “commutable between lots.”

The interpretation of the combined assessment is illustrated in principle in Fig. 2 and for 3 participants using different POC INR methods in Fig. 4. The best-case scenario is when both results are within the quality specifications (field A), implying that the POC method has an acceptable bias and that the participant performs the test in concordance with the manufacturer’s procedure. Field I indicate a worst-case scenario in which a primary care laboratory overestimates the results compared with others using the same POC method, and in addition applies a POC method with an unacceptable positive bias. This laboratory will report patient results that are greatly underestimated. Field F, on the other hand, illustrates a situation in which a laboratory “compensates” for a positive method bias with a negative performance deviation. This laboratory underestimates the result compared to others with the same method, but can get a nearly “correct” result for the patient. Laboratories in both these circumstances will receive the message that they should change to a better method and improve their performance.

More than 90% of the participants achieved EQA INR results within the quality specification, whereas only 2 of the 3 POC INR methods fulfilled the quality specification for method bias. These data indicate that more effort should be put into method improvement, or advising against using “poor” methods, rather than on addressing participant performance.

Despite application of the proposed model twice with POC INR, further evaluation of the model is needed. For example, a user evaluation among the participants should be conducted to examine whether or not they understand this kind of feedback. For some participants, interpreting the results of this model can be difficult, and supplying feedback regarding the method bias simply as plain text might be more easily understood than Fig. 2. However, the proposed model should be very useful for EQA organizations and for laboratory consultants who give advice to the participants. It will also be of value for the continuous monitoring of method harmonization (4).

In conclusion, we propose an EQA model that combines assessment of both method bias and single-participant performance. This model should be easy to carry out on a regular basis and can be used for constituents where the EQA control material is noncommutable.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: The authors thank all the expert primary health-care centers for participating in this study, and the personnel at the laboratory at Haraldsplass Deaconess Hospital for analyzing the native patient samples on the comparison method.

References

7. CLSI. Quality management: approaches to reducing errors at the point of care; approved guide-

8 Clinical Chemistry 59:2 (2013)


