What have we learned from coagulation laboratory participation in external quality programs?

Kristi J. Smock | Karen A. Moser

Department of Pathology, University of Utah Health Sciences Center and ARUP Laboratories, Salt Lake City, Utah

*Correspondence
Kristi J. Smock, Department of Pathology, University of Utah Health Sciences Center and ARUP Laboratories, Salt Lake City, UT. Email: kristi.smock@aruplab.com

Abstract
Coagulation laboratory external quality assurance (EQA) programs provide information that satisfies regulatory requirements and improves laboratory quality, patient care, and safety. In addition to the value to individual laboratories, data from EQA programs benefits the laboratory community in multiple ways by providing a snapshot of laboratory practice and summarizing the performance of various methods in identifying normal and abnormal specimens and the effects of therapies or interfering substances. This review article aims to summarize and provide examples of some of the important lessons learned from coagulation EQA data, including issues related to non-standardization, imprecision, and the effects of interfering substances.

KEYWORDS
Coagulation, D-dimer, direct oral anticoagulants, external quality assurance, factor assays

1 | INTRODUCTION

Participation in external quality assurance (EQA) programs, where laboratories test unknown specimens, also known as proficiency testing, is essential for all clinical laboratories, including hemostasis/thrombosis (coagulation) laboratories. Although participation is generally required to satisfy regulatory and accreditation requirements, comparison of laboratory data to the summary data returned by the EQA organization provides value far beyond simply "checking a box" for participation. For instance, an individual laboratory will learn how its results compare to peers using the same instrument/reagent combinations for a given analyte and also how its results compare to different testing systems, information which improves quality, patient care, and safety. In some EQA programs and for some analytes, performance is formally graded with a clear pass or fail designation that may be reported to regulatory agencies, while other times the laboratory will evaluate its own performance using local policies. The overall performance trends yield extensive information and give a snapshot of laboratory practice that includes availability of testing for certain conditions, the spectrum and frequency of methods being used, peer-based (or sometimes accuracy-based) performance of individual methods (both numeric results and qualitative interpretation), and the effects of interfering substances across multiple methods. In addition, granular information about laboratory methods and practices is often collected using questionnaires that accompany the testing surveys. This review article will focus on valuable lessons that have been learned by coagulation laboratory participation in EQA programs with an emphasis on how the lessons can be applied to improve the quality of laboratory coagulation results.

Although many coagulation EQA programs have been in place for decades, most beginning with evaluation of routine coagulation tests such as prothrombin time (PT), activated partial thromboplastin time (aPTT), and fibrinogen, current programs are much more comprehensive. Analytes include routine clotting times, factor assays, inherited and acquired thrombophilic risk factors, fibrinolytic system components, anticoagulant monitoring, some platelet function tests, viscoelastic tests, and others. Examples of recent additions to EQA programs include testing for ADAMTS13 activity and autoantibodies, chromogenic factor assays, and measurement of direct oral anticoagulants (DOACS). EQA specimens are designed to include normal and abnormal conditions, encompassing disease states, therapies, or substances expected to interfere with testing. There are often some gaps in EQA offerings due to the constant and at times rapid evolution in medical practice guidelines, laboratory technology, and approved medications.
as well as the difficulty in evaluating certain tests in an EQA setting in a way representative of patient testing (e.g., platelet function testing). In recent analyses of coagulation EQA data, some of the most important lessons learned involve the effects of lack of standardization or harmonization for many coagulation methods, which may be problematic when non-standardized tests and/or reporting units are used in clinical algorithms and guidelines, the effects of inaccuracy or imprecision on disease diagnosis and classification, which may be related to non-standardization, and the effects of interfering substances on coagulation testing, which may vary significantly depending on the laboratory method. Although there are many informative publications highlighting various aspects of EQA performance, and a number of very active EQA organizations worldwide, the specific examples we have selected to discuss represent only a sampling that have had direct relevance to our laboratory practice.

2 | D-DIMER

One assay where EQA program data has illustrated both lack of assay standardization and variability in result reporting is D-dimer. Many different D-dimer kits are commercially available. Additionally, several different magnitude units (e.g., ng/mL, μg/mL, and μg/L) and two different unit types (D-dimer unit (DDU) and fibrinogen equivalent unit (FEU)) are regularly used, and a recent international survey identified 28 combinations of unit magnitude and type reported by survey participants.4,6 Previously published EQA data from the College of American Pathologists (CAP) have demonstrated confusion resulting from different reporting units.4 First, the CAP EQA data showed that some laboratories convert between unit types.4 For example, a laboratory’s assay kit may recommend reporting results in DDU but a laboratory will choose instead to convert to FEU using the predictable mathematical relationship between the two units (1 ng/mL DDU = 2 ng/mL FEU). In 2004 CAP data, 34% of laboratories converted from one unit type to another, based on calculation from reported results. Following educational initiatives by the CAP Coagulation Resource Committee, including an informational letter to D-dimer assay manufacturers emphasizing the importance of clear instructions for unit reporting in package inserts, in 2011, 12.7% of laboratories converted from one unit type to another, based on self-report.4 Further, evaluation of D-dimer results by the CAP identified a bimodal distribution of results in 2004, 2007, and 2011, suggesting that laboratory confusion about the type of units may lead to incorrect result reporting.4 Despite the predictable mathematical relationship between DDU and FEU, converting between units is not a recommended practice, due to the risk of error.7,8 Although the variability in D-dimer reporting is apparent from reviewing the instructions for use for commercially available kits, D-dimer survey data reveal the scope of problematic result reporting.

Further complicating D-dimer testing is the lack of standardization or harmonization between different D-dimer assay kits. There is currently not an international standard for D-dimer assay calibration.9 Development of such a standard has been attempted in the past, but was unsuccessful, likely due to the different D-dimer epitopes recognized by varied antibodies in different manufacturers’ kits.6,9,10 2007 CAP data show wide variation between method-specific peer group means, ranging from 295 to 2108 ng/mL FEU in a slightly elevated D-dimer sample (with an all method mean of 1568 ng/mL FEU) and ranging from 470 to 10 150 ng/mL FEU in a moderately elevated D-dimer sample (with an all method mean of 3772 ng/mL FEU).4 College of American Pathologists data also indicate significant differences in D-dimer assay imprecision, reporting an all method coefficient of variation (CV) of 25.5% (range of method specific CVs 6.3%-30.5%) for a slightly elevated D-dimer sample and all method CV of 22.8% (range of method specific CVs 4.8%-25%) for a moderately elevated D-dimer sample.4 These data illustrate the effects of lack of assay standardization and assay imprecision in a clear and understandable way.

The lack of D-dimer assay standardization and confusion surrounding D-dimer reporting exemplified by EQA data leads to problems describing D-dimer assay results in the medical literature. We are aware of several examples of D-dimer use in clinical diagnostic algorithms and/or decision rules where application of the proposed algorithm or rule may be hampered if users cannot identify the correct D-dimer unit or if the local laboratory’s D-dimer assay cannot be used in the proposed tool. For example, clinical guidelines published by the American College of Physicians (ACP) in 2015 challenged laboratories to decide whether age adjusted D-dimer cutoffs used in protocols for exclusion of venous thromboembolism are safe and appropriate for use in their institutions, a decision which includes determination of whether or not their D-dimer assays have sufficient data to support the use of age-adjusted cutoffs.5,11 Unfortunately, the unit type (DDU versus FEU) of D-dimer results was not clearly stated in the initial ACP guidance paper, but was subsequently corrected to indicate a unit type of FEU11. Failure to communicate the D-dimer unit type is common in published literature in this area, including in a recent proposed guideline update for diagnosis of periprosthetic joint infections that does not specify either the D-dimer assay used by study sites or the D-dimer unit type.12 The HERDOO2 clinical decision rule for anticoagulation duration in women with unprovoked venous thromboembolism was developed using a single D-dimer assay.13 However, subsequent studies suggest that not every D-dimer assay can be used in the HERDOO2 rule, a clinical consequence of the significant D-dimer assay variability previously observed in EQA data.14

3 | FACTOR ACTIVITY AND INHIBITOR ASSAYS

In another example, laboratory performance in testing for coagulation factor activities is critical for diagnosis and severity classification of hemophilia A (factor VIII (FVIII) deficiency), hemophilia B (factor IX (FIX) deficiency), von Willebrand disease, rare inherited factor deficiencies, acquired deficiencies, and to monitor therapy
with clotting factor replacement.\textsuperscript{15} The severity of hemophilia is based on FVIII or FIX activity with <1% classified as severe disease and moderate- or mild- disease characterized by 1%-5% and 6%-40%, respectively.\textsuperscript{15,16}

The earliest iterations of the CAP EQA program in hemostasis included FVIII and in analyses of CAP data from the 1970’s and 1980’s, FVIII activity precision was classified as very low (using method-specific CVs) for FVIII between 1% and 20%, low to intermediate for specimens containing normal FVIII levels, and the method-specific CVs did not improve over time.\textsuperscript{1,17-23} In the latest CAP publication from 1987, a large number of laboratories (between 413 and 577) were challenged with three plasma specimens at different FVIII levels and the following all-method CVs (with precision classified as very low to low) were demonstrated: 84.6% (specimen containing 2% FVIII activity), 29% (specimen containing 27% FVIII activity), and 27% (specimen containing 60% FVIII activity).\textsuperscript{23} Overall, these older data indicated significant between- and within- method variability. Similar low precision was also noted in CAP studies from the 1980’s for FIX and factor XI (FXI) assays.\textsuperscript{1,24,25}

The North American Specialized Coagulation Laboratory Association (NASCOLA) proficiency testing program is part of the External Quality Control of Diagnostic Assays and Tests (ECAT) Foundation EQA program. NASCOLA data from 2004 indicated CVs of 19% for a specimen with low FVIII (35% of normal) and 12% for normal FVIII (100% of normal) from a group of approximately 30 laboratories.\textsuperscript{26} In 2013, NASCOLA published an analysis of the North American data for factor VII (FVII) assays for 24 specimens tested by over 50 laboratories.\textsuperscript{27} Their findings indicate better between-laboratory precision for normal specimens (CV 10.7%) versus abnormal specimens containing FVII <20% (CV 33%) and within-method CVs that were similar to the between-method CVs and to each other. Similarly, in a NASCOLA study of FXI activity testing, between laboratory precision was acceptable for samples with normal and mildly reduced FXI activity and the highest CV was observed in a specimen with markedly low FXI activity.\textsuperscript{28}

In a summary of data from the Royal College of Pathologists of Australasia Quality Assurance Programs (RCPAQAP), ECAT, and UK National External Quality Assessment Service (NEQAS) EQA programs for 2011 and 2012, FVIII and FIX assays demonstrated inter-laboratory CVs from 15% to 25%, with CVs increasing as factor level decreased.\textsuperscript{15} The data also show occasional outliers, including occasional extreme outliers, that indicate inaccurate test results, and could result in misclassification of disease severity if similar inaccurate values were obtained in clinical practice.\textsuperscript{15}

Questionnaire data has revealed many differences between laboratories that may influence factor assay accuracy and precision. These include patient plasma collection and handling and variables such as instrumentation, calibrator and reagent sources, and number and magnitude of patient dilutions tested. In the NASCOLA FVII study, an example from 2008 data showed use of 8-10 thromboplastin reagents, 9 calibrators, 9 FVII-deficient plasmas, and 8-10 analyzers in various combinations by the participants and some differences were seen related to thromboplastin reagents and calibrators.\textsuperscript{27} In the NASCOLA FXI study, a significant difference was not seen between results when test-related variables were considered individually.\textsuperscript{28} The local reference interval also impacts classification of a result as normal or abnormal. The NASCOLA studies have highlighted differences in this regard.\textsuperscript{27,28} For example, the same numeric value could be classified as normal by one laboratory and abnormal by another when values are near the lower end of the reference interval, which could affect the clinical impression.

Identification and titering of coagulation factor inhibitors, which greatly impacts hemophilia treatment using clotting factor concentrates, is generally performed with a Bethesda assay, which involves determining the recovery of factor activity from a mixture of patient plasma dilutions containing the antibody and normal pooled plasma and using the residual factor activity to determine antibody titer. Thus, Bethesda assay performance is dependent on a number of variables, including the underlying factor assay used to determine residual factor activity. The Factor VIII and IX Subcommittee of the Scientific and Standardization Committee (SSC) of the International Society on Thrombosis and Haemostasis (ISTH) indicates that relevant inhibitors are persistent in two specimens collected on two separate occasions in a 1-4 week period and should have a level of ≥0.6 Bethesda units per mL using the Nijmegen modification of the Bethesda assay.\textsuperscript{16} Further, the definition of low- and high- titer inhibitors is determined by antibody titers ≤5 or >5 BU/mL, respectively. Low and high-titer inhibitors have different treatment approaches, and Bethesda titers are followed throughout treatment.\textsuperscript{15,16,26} It has been shown that laboratories have variable lower limits of detection for the Bethesda assay (ranging from 0 to 1 Bethesda units in a NASCOLA publication) and, additionally, that methods that are hybrids of the classical Bethesda and the Nijmegen modification are most commonly used in North American laboratories.\textsuperscript{26} NASCOLA has also reported high between-laboratory CVs ranging from 30% to 42% for low-level FVIII inhibitor titers and even higher between-laboratory CVs (56%) for high-titer FVIII inhibitors, indicating that CVs for inhibitor titer are even larger than those for factor assays, likely due to the increased complexity of inhibitor assays.\textsuperscript{26} Similarly, a study published by ECAT including data from 100 to 170 laboratories who received challenges between 2006 and 2008 found overall between-laboratory variation ranging from 28% to 52% and with slightly lower variability for the Nijmegen assay protocol.\textsuperscript{29} RCPAQAP Haematology and UK NEQAS EQA have also reported high inter-laboratory CVs (often >30%) for inhibitor testing, occasional false-negatives when an inhibitor is present, and false-positive low titer inhibitors when an inhibitor is absent, and decreased variability with Nijmegen protocols compared to standard Bethesda assays.\textsuperscript{15}

The findings described above highlight the existing variability in factor assay and factor inhibitor methods, a need for greater factor assay precision in specimens with low factor activity, and greater precision in tests for inhibitor titer. Although guidelines for patient diagnosis and management utilize definitions based on laboratory values, imprecision of the assays at clinically important decision levels prevents the application of these guidelines in an optimal and consistent way. In addition, the landscape is becoming increasingly complex due
to the recent development of several extended half-life factor replacement products (modified recombinant FVIII and FIX concentrates). Some of these products show a significant difference in recovery from labeled potency, including over- or under-recovery, depending upon the concentrate and assay methodology, and necessitating use of chromogenic factor assays rather than one-stage clot-based assays for monitoring in some cases. However, the one-stage clotting assays remain the most frequently used worldwide. Although the differences in factor assay recovery of these new products have mainly been reported in field studies and by manufacturers, rather than studies performed within EQA programs, these programs could provide a venue for evaluating the real-world performance of factor assays in measuring these products in a large number of laboratories and some studies within EQA programs are currently underway. Additional changes in laboratory practice are also arising due to the development of non-factor therapies, such as emicizumab, a bispecific antibody that bridges FIXa and FX and replaces FVIII function which is now approved for treatment of hemophilia A with or without inhibitors. There are currently no approved laboratory assays to monitor therapy with these novel assays will also need to be included in EQA programs.

4 | DIRECT ORAL ANTICOAGULANTS

Standardization, accuracy, and precision issues are also important for the management of anticoagulant therapy. Laboratory variability in anticoagulant monitoring (e.g., heparin and warfarin) is a well-known phenomenon. However, less is known about how tests perform in the measurement of the DOACS since the drugs do not require routine monitoring and access to testing is limited. For instance, in the United States, there are currently no FDA-cleared assays for measuring DOACS (such as dabigatran, rivaroxaban, apixaban, and edoxaban), and data from a 2015 NASCOLA survey indicated that approximately half of the special coagulation laboratories that participated did not offer tests for dabigatran, rivaroxaban, or apixaban. Nonetheless, measurement is desired in certain situations such as overdose, differentiation of treatment failure from non-compliance, life-threatening hemorrhage with potential need for drug reversal, or need for emergent surgery or thrombolytic therapy, and guidance has also been issued from several prominent scientific societies recommending more routine evaluation of drug-specific peak and trough levels in certain patient populations, such as those with impaired renal function or high body mass index (BMI), who may have increased or decreased drug exposure, respectively.

The activities of EQA programs have provided some information regarding test performance for measurement of DOACS. A publication of ECAT and NASCOLA data acquired in 2013 described the results of predominantly European and North American laboratories that were challenged with two specimens containing dabigatran and two specimens containing rivaroxaban. These international results showed a variety of assays and calibrators used to measure the drugs. For dabigatran, approximately 96% of results received a passing score for their quantitative results when comparing individual lab results to mean values. Statistically significant differences were not seen among different reagents or instruments but were seen between commercial versus homemade calibrators. When between-laboratory precision was assessed, the two most commonly used reagents demonstrated intermediate precision (CVs in the 11%-20% range), the three least commonly used methods met criteria for poor precision (CV > 20%), and some reagent-instrument differences were observed. Similar to dabigatran, 96% of results received a passing score for rivaroxaban. For rivaroxaban, one reagent tended to generate lower results for a 100 ng/mL specimen compared to four other common reagents, but this was not seen in a specimen containing a higher rivaroxaban concentration (300 ng/mL). Differences were not seen among calibrators or instruments. For the most common methods, precision was classified as intermediate for the specimen with a lower rivaroxaban concentration but was somewhat better for a 300 ng/mL specimen. As described by the study authors, the between-laboratory variability observed could have clinical implications since the results obtained by different laboratories on the same dabigatran specimen could be different enough to be interpreted as consistent with either typical peak values or typical trough values. A similar example was also provided based on the rivaroxaban precision data. The authors conclude that an international standard available to manufacturers for calibration of their drug calibrators may help to diminish assay variability.

Inter-laboratory variability in DOAC measurement has also recently been reported in EQA data from the Italian Federation of Thrombosis Centers and RCPAQAP. In Italy, quantitative results from dedicated tests for DOACS were available from approximately one-third of the 235 participating laboratories, highlighting the limited availability of this testing. The remainder of the laboratories reported only the results from their routine clotting times for specimens containing DOACS. Relatively minimal between-laboratory variability was observed with overall CVs of 8.7% for dabigatran, 8.4% for rivaroxaban, and 10.3% for apixaban measurement. It is unclear why the CVs observed in this study indicate less assay variability than was seen in the ECAT/NASCOLA study. Results from RCPAQAP included 70 laboratories evaluating samples containing dabigatran and 44 laboratories evaluating samples that contained either rivaroxaban or apixaban. There were two main assay groups for measurement of dabigatran included in the RCPAQAP. Hemoctol DTI/laboratory developed dilute thrombin time (dTT) methods and ecarin clotting time (ECT). The dTT methods demonstrated CVs between 10% and 33%, while ECT methods showed a higher range of CVs (32%-119%). This study also indicated high reproducibility between the dTT methods, with similar results obtained when similar dabigatran-containing samples were measured in two different surveys. For anti-Xa activity assays using calibrators for rivaroxaban or apixaban, RCPAQAP reported good correlation between the
expected drug level and the level reported by anti-Xa activity for a
series of 8 samples with concentrations ranging between 0 ng/mL
and approximately 800 ng/mL. The RCPAQAP described increased
variation at very high rivaroxaban levels, as well as insensitivity
to very low levels of both rivaroxaban and apixaban (below about
20-30 ng/mL). Additional study of DOAC assay performance by
other EQA programs is underway and will likely yield further useful
information. Ultimately, the contributions of EQA data to the litera-
ture provide evidence that regulatory clearance for DOAC assays is
needed to expand test availability and encourage standardization at
the manufacturer level that could help minimize potentially clinically
significant assay variability.

Data from EQA programs addressing the effects of DOACS on
other special coagulation assays (not those used for monitoring
DOACS) are also emerging. Although many single and multicenter
studies have described the performance of special hemostasis as-
says in plasma samples containing DOACS, few published studies
addressing the performance of these assays in EQA programs exist
at the time of this review.

A review of protein S (PS) testing performance in North American
clinical laboratories from 2010 to 2012 through the NASCOLA
program included evaluation of one normal pooled plasma sample
spiked with 200 ng/mL rivaroxaban.38 The all-method mean PS ac-
tivity overestimated functional PS as compared with free PS antigen
assays (mean PS activity 112%, mean free PS antigen 80%). In addi-
tion, PT- and Russell Viper Venom Time (RVVT)-based PS activity as-
says showed higher mean PS activity results than the aPTT-based PS
activity assays, suggesting that PT- and RVVT-based PS activity as-
says are affected by rivaroxaban interference to a greater degree.38

The Belgian national External Quality Assessment Scheme
(EQAS) has published results of two studies in which it sent normal
pooled plasma samples spiked with differing concentrations of dab-
igatran, rivaroxaban, and apixaban to approximately 190 participat-
ing laboratories for measurement of routine coagulation assays as
well as antithrombin (AT) activity.40,41 Laboratories that used AT ac-
tivity assays based on inhibition of thrombin tended to overestimate
antithrombin activity in the presence of 250 μg/L dabigatran.40,41
Likewise, laboratories using AT activity assays based on inhibition of
factor Xa tended to overestimate antithrombin activity in the pres-
ence of rivaroxaban (best seen at 290 μg/L) and apixaban (seen at all
concentrations tested, but best seen at 225 ng/mL).40,41

In addition, the RCPAQAP studies mentioned above also in-
cluded information on the effects of dabigatran, rivaroxaban, and
apixaban on special hemostasis assays.37,38 The RCPAQAP data
demonstrated no effect of dabigatran, rivaroxaban, or apixaban on
chromogenic protein C activity or free protein S antigen. Particularly
with drug concentrations >100 ng/mL, dabigatran caused falsely in-
creased anti-IIa based antithrombin activity, while rivaroxaban and
apixaban produced falsely increased anti-Xa based antithrombin
activity. All three drugs showed concentration-dependent false de-
creases in APTT-based factor activities, and rivaroxaban and apix-
aban also demonstrated concentration-dependent false increases
in PT-based factor activities (with rivaroxaban concentrations
>100 ng/mL showing a greater effect on the PT-based activities
than apixaban). Dabigatran and rivaroxaban showed potential
false positive dilute Russell viper venom time (dRVVT) results for
lupus anticoagulant, while apixaban prolonged both the screening
and confirm dRVVT to a similar degree and did not result in a false
positive dRVVT ratio in the RCPAQAP study. Finally, the RCPAQAP
assessed laboratory performance of factor VIII inhibitor assays in
the presence of 800 ng/mL dabigatran and 638 ng/mL rivaroxaban.
Both drugs resulted in false positive inhibitor levels, with dabiga-
tran producing higher Bethesda unit (BU) values than rivaroxaban
dabigatran range 1.0-55.6 BU, rivaroxaban median 0.5 BU and
range 0-3 BU).37,38 The combined findings from multiple EQA pro-
gram publications support other single and multicenter laboratory
studies of the interfering effects of DOACS on special coagulation
tests and confirm that the observed patterns of interference hold
true in larger peer groups across many different laboratories.42-44
Additional EQA studies are in progress. In addition, summary tables
provided in review articles provide a comprehensive discussion of
observed DOAC assay interference from studies performed outside
of EQA settings, and interested readers are referred to such tables
for additional information.45

5 | CONCLUSION

In conclusion, we have highlighted several examples of lessons
learned from EQA data that relate to assay non-standardization,
assay imprecision, and assay interferences. These lessons highlight
some of the benefits of EQA participation previously described,
namely improved laboratory practice to improve patient care and
safety, characterization of test precision using multiple laboratory
methods, correlation of specific test variables with accuracy and
precision, and description of interfering substance effects on mul-
tiple methods.1 Continued description of test performance across
varied instrument and reagent combinations by EQA programs will
continue to improve laboratory practice. Although not a specific
focus of this review, it should also be noted that there are many
initiatives in progress attempting to internationally harmonize EQA,
since different programs analyze data differently, and these efforts
will further improve the utility of these programs.2

ACKNOWLEDGEMENTS

Both authors Smock and Moser wrote significant portions of the
manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ORCID

Kristi J. Smock https://orcid.org/0000-0003-0304-6461
REFERENCES


How to cite this article: Smock KJ, Moser KA. What have we learned from coagulation laboratory participation in external quality programs?. *Int J Lab Hematol*. 2019;41(Suppl. 1):49-55. https://doi.org/10.1111/ijlh.12998