# Duke Virology Quality Assurance


## Quantitative HIV-1 RNA PT Program
## Participation Requirements and Scoring Procedures

Thomas Denny MSc, M.Phil., Principal Investigator
Raul Louzao, MPA Laboratory Operations Director
Miranda Carper, PhD, VQA Program Manager
Darin Weed, PhD, VQA Program Manager
Salvatore Scianna, QCM Director

# Contents

# VQA Program Quantitative HIV-1 RNA Proficiency Testing Requirements

## Introduction to Participation Requirements

The National Institute of Allergy and Infectious Diseases (NIAID), Division of AIDS (DAIDS) Virology Quality Assurance (VQA) Program provides quality assurance and proficiency testing (PT) to labs that are performing virologic-based assays for HIV and other pathogens in support of NIAID-funded clinical trials. All laboratories that intend to do Quantitative HIV-1 RNA testing for NIAID-funded clinical trials must have an Acceptable performance rating in the VQA Quantitative HIV-1 RNA proficiency testing program.

Participation in the real-time testing phase of the program can only occur after the successful completion of a prequalification panel or assay validation. VQA HIV-1 RNA proficiency testing panels (5 samples) are to be run every three months and must be assayed in a single batch in the month in which the panel is due. The following sections detail the requirements to obtain and maintain an Acceptable performance rating to participate in quantitative HIV-1 RNA proficiency testing for NIH-funded protocols.

## Proficiency Testing Overview:

Beginning January 2023 new scoring rules will be in effect (please see **VQA Program Quantitative HIV-1 RNA Proficiency Testing Scoring Procedure** below for detailed information on scoring procedures). Under the newly approved scoring system sites will now receive a score of either Satisfactory or Unsatisfactory for datasets submitted for each PT. Additionally, there will be two performance rating categories, Acceptable or Not Acceptable. A laboratory must have a Satisfactory grade for 2 out of the 3 past PTs to have an Acceptable performance rating.

The performance ratings are listed in **Table 1** (see the **VQA Program Quantitative HIV-1 RNA Proficiency Testing Scoring Procedure** section below for detailed information on scoring):

**Table 1: VQA Performance Ratings**

| Assessment | PR | PT Score |
|---|---|---|
| Acceptable | A | SSS, SUS, USS, SSU |
| Not Acceptable | NA | UUU, SUU, USU, UUS |

PR: Performance Rating
A: Acceptable; NA: Not Acceptable
S: Satisfactory; U: Unsatisfactory

Laboratories will be shipped four panels once a year. Each panel consists of 5 samples and will be named using the following format:

- RNAYYYY_MM.01-05A-E
  - RNA = PT program
  - YYYY = year of the PT
  - MM = month of PT testing
  - 01-05 = sample number
  - A-E = panel configuration

Laboratories will test a 5-member panel in a single assay every three months (February, May, August, and November). Sites are no longer required to run a VQA200 control with their PT.

The VQA Biostatisticians will analyze data for each PT. Each analysis will include the current and three previous rounds of PT data for a total of 20 samples. The analysis for each PT includes 20

samples because it provides sufficient statistical power. Scores for each PT will be provided after each round of testing in a report that will be distributed to the participating laboratories.

Scores for each round of testing will be derived from an analysis of precision, accuracy, specificity, sensitivity, and assay validity (see **Table 2**). Precision assessments will be based on the combination of inter-assay and intra-assay variations to better monitor the variability associated with real-time testing. Expected values for total variation were derived empirically from VQA proficiency testing data, and the distribution of total assay variation was determined using Monte Carlo simulations. Proficiency testing data will be scored for deviations in total assay variability based on these cut points. Comparing the average $\log_{10}$ recoveries in a laboratory with the median $\log_{10}$ recovery across all laboratories will assess accuracy. Median values will be used instead of mean values to eliminate effects of outliers. Sensitivity will be determined for each assay by testing specimens at or near the limit of detection. All assays are challenged for sensitivity using samples with a nominal value of 50 copies/mL and for specificity using blinded samples that do not contain HIV.

**All invalid proficiency panel samples and assays must be repeated prior to data submission**. Laboratories will also be evaluated on transcriptional and computational errors. Laboratories are encouraged to submit their raw data file.

**Table 2: Quantitative HIV-1 RNA Proficiency Criteria and Targets**

| Criteria | Description | Target |
|---|---|---|
| Precision | Total assay standard deviation based on the ratio of the estimated viral load (laboratory value) to the expected viral load. | Less than the 95th percentile |
| Accuracy | Measure of how closely a laboratory estimate of viral load is to the expected median $\log_{10}$ value | Less than 3 standard errors from the pooled laboratory median |
| Sensitivity | Number of false negative results obtained from replicates at the limit of detection for the assay in question | One negative result of 50 copies/ml replicate is allowed |
| Specificity | Monitor the results obtained on samples that do not contain HIV | All 0 copies/ml replicates must be not detected |
| Assay Validity | It is expected that all results should meet established criteria for assay validity, which varies among kits. Assay validity is based on manufacturer-defined controls included in the run. | No invalid assay runs |

## VQA Participation Requirements

New Laboratory Qualification Testing

Participation in any VQA proficiency testing program must be approved by the VQA Contracting Officer Representative (COR, aka Project Officer). In order to be considered for participation in the VQA program, a new laboratory must be doing quantitative HIV-1 RNA testing for a NIH-funded program. Once approved, the new laboratory will need to complete an application for participation in the program, which will outline the programs of interest and provide laboratory contact

information. A VQA laboratory is identified by a Harmonic ID (HID), Laboratory Data Management System (LDMS), or VQA assigned number.

Prior to enrollment into the real-time HIV RNA proficiency testing program, all new laboratories will need to test a 20-member prequalification panel containing coded plasma specimens or perform an assay validation that consists of VQA controls and clinical samples. The prequalification panel and the validation data will be submitted to the VQA via Duke Box or the VQA Web Application. Please contact the VQA with any requests to enroll into the Quantitative HIV-1 RNA PT program via vqa-rna@duke.edu. Laboratories that test a prequalification panel will need a score of Satisfactory prior to entering into the HIV RNA proficiency testing program. A score of Unsatisfactory will require the laboratory to run another 20-member panel prior to advancing. A second score of Unsatisfactory will require intervention by the VQA prior to any additional testing.

A new laboratory must obtain an Acceptable performance rating prior to performing any protocol testing. After the successful completion of a prequalification panel or an assay validation using VQA controls, a site will need to receive a score of Satisfactory on two out of three of the most recent five-member panels.

Panels may be tested during regular rounds of proficiency testing, which occurs four times per year, or they may be Fast-Tracked using historical panels that have been pre-tested in the field. The normal track or fast track options must be begun within one year of the successful completion of the prequalification panel or assay validation. If PT testing is postponed beyond one year, then another 20-member prequalification panel must be passed before participation in the VQA Quantitative HIV-1 RNA PT program.

**The final determination of whether or not a laboratory may perform protocol testing is at the discretion of the program leadership for that laboratory, not the VQA.**

New Assay Enrollment and Qualification

A laboratory may request to:

- Qualify additional assays for protocol testing
- Switch platforms
- Validate additional instruments for qualified assays

All new assays and instruments enrolled in the VQA program need to be qualified following the procedures outlined in the above section, **New Laboratory Qualification Testing.** The VQA permits the use of two different assays; laboratories must obtain special approval from the VQA if they wish to become qualified using three or more different assays. Each assay added to the VQA Program will be qualified separately and the rules for achieving an Acceptable performance rating applies to all assays used within a laboratory. If a laboratory uses a manual extraction method as a backup for their automated extraction instrument, then the laboratory will need to demonstrate ongoing proficiency testing using both extraction methods.

The VQA can assist in assay and instrument validations. Please contact vqa-rna@duke.edu for more information about assay requirements or to request a validation plan and data submission template.

### Maintaining an Acceptable Performance Rating

To maintain an **Acceptable performance rating** a laboratory must do the following:

- Participate in the routine proficiency testing program as scheduled by the VQA
- A laboratory must have a Satisfactory score for 2 out of the 3 most recent rounds of proficiency testing.

Extensions will no longer be granted to sites. PT due dates will be adjusted as needed if the shipment of a panel is delayed. If a lab does not submit data by the PT due date the lab will receive a score of Unsatisfactory. However, a laboratory may request a PT exemption if they are unable to participate in proficiency testing due to circumstances out of their control. All exemption requests must be approved by the VQA. Acceptable reasons to temporarily be exempt from proficiency testing include:

1. Force Majeure (hurricane, tornado, flood, fire, etc.)
2. Laboratory closures for emergency circumstances (fire, pandemic, radiation leak, flood, electric issues, *etc*.)
3. Supply chain issues with vendor
4. Broken instrument
5. Government shutdown or political unrest

Exemption requests should be made to vqa-rna@duke.edu prior to the start and/or close of the PT. A laboratory may receive a score of Unsatisfactory if the reason for not submitting results does not meet one of the acceptable criteria for exemption. **A laboratory must contact their network leadership if they are exempted from a round of VQA Quantitative HIV-1 RNA proficiency testing.**

### Repeating Testing of PT Panels

During the real-time testing phase of the program, problems may arise that require a laboratory to repeat a given five-member panel. Since problems that affect one five-member panel may affect future analyses, each laboratory is given the option of repeating any five-member panel ("B" panel configurations are used for repeating five-member panels). If a laboratory receives a score of Unsatisfactory, the lab will receive an Investigation Report (IR) that is to be reviewed and filled out by the laboratory. The IR form helps guide the lab in determining the root cause of the score for that PT. If a laboratory receives a score of Unsatisfactory for a given round, they must repeat the five-member panel. The VQA will contact a laboratory regarding the option of repeating a panel if the laboratory was issued a Potential Issue Alert (PIA). PIAs are informal notices issued to any lab that has accuracy statistics that indicate average recovery is at least three but less than four standard errors from the expected value; standard deviation (SD) between the 95th and 99th percentile; and one invalid sample. Any laboratory with a score of Satisfactory that wishes to repeat a panel must contact the VQA by emailing vqa-rna@duke.edu in order to receive a repeat panel. All laboratories that receive a score of Unsatisfactory will automatically receive a repeat panel.

A five-member panel may only be repeated once, and may only be repeated during the round in which the error occurred. The repeat panel data, as well as the next five-member panel data, will be due by the deadline for the next round of testing unless the VQA instructs otherwise. **Repeat ("B") panel samples must not be analyzed in the same run as any other VQA samples to enable proper precision scoring**. **A laboratory will not receive a separate analysis for the "B" round testing, but the new data will replace the problematic data in future analyses.**

If a laboratory opts not to repeat samples, then problems that were noted may carry over to subsequent analyses. As necessary, the VQA will ship an extra 5-member panel to a laboratory that fails a round of testing. Ongoing scoring problems may require a laboratory to undergo requalification (successfully pass a 20-member panel) in order to maintain eligibility for protocol testing (see the **Requalification** section below).

## Withdrawal/Removal

A laboratory may voluntarily withdraw from the VQA HIV RNA proficiency testing program at any time. A laboratory that has not participated in the VQA HIV RNA proficiency testing program for more than 12 consecutive months will automatically be removed from the program.

## Requalification

Labs Placed on Hold or Withdrawn from VQA Program

If a laboratory wishes to re-enter the VQA HIV RNA proficiency testing program subsequent to removal or has not participated in longer than 12 months, that laboratory will need to be requalified as a new laboratory (see section above).

Following Ongoing Proficiency Testing Problems

A laboratory that is having ongoing RNA proficiency panel testing problems and is not maintaining an Acceptable performance rating may be asked to undergo new laboratory requalification. In this case, a laboratory will be provided a 20-member panel that will be tested in a single assay. A laboratory must achieve a score of Satisfactory on this requalification panel in order to resume real-time testing. Additional panels may be needed in order to achieve this score. A laboratory will need to contact the VQA via vqa-rna@duke.edu in order to request requalification panels.

For requalification, the 20-member panel will be scored in the same manner as a qualification panel for a new laboratory. Assay sensitivity, specificity, accuracy and intra-assay precision statistics will be applied. As a laboratory participates in proficiency testing, 5 qualification samples will be removed in sequential order for each new 5-member PT panel the laboratory analyzes. As a lab participates in PTs, data from the 20-member panel will be used until four 5-member panels have been completed. Each new 5-member panel will be added, and the last 5-member panel data will be removed, providing the cumulative scoring described in the **VQA Program Quantitative HIV-1 RNA Proficiency Testing Scoring Procedure** section below**.**

Precision scoring for the 20-member requalification panel will only include the intra-assay component of variation. Once a newly submitted panel is added to the requalification process, total assay precision statistics (including both inter- and intra-assay components of variation) will replace intra-assay precision statistics. An Acceptable performance rating will occur after a laboratory achieves a score of Satisfactory on two out of three of the most recent five-member panels.

## Appeals

The VQA recommends scoring for proficiency panels based on the criteria defined for the program. The VQA Advisory Board (VQAAB) then reviews the scoring for each round of testing for labs that had a decline in status or had a problem that does not follow the VQA PT rules (the laboratory identities are blinded for this process). Any laboratory may appeal the score on a proficiency panel by submitting a letter or email to William Meyer III, Chair of the VQAAB (william.a.meyer@questdiagnostics.com). All appeals will be reviewed by the VQAAB to determine if a change in scoring is indicated. Laboratories will be notified of the outcome of all appeals.

## Practice Panels

Any NIAID approved laboratory may obtain retrospective HIV RNA proficiency panels prior to enrollment in the program, or at any time during their participation in the VQA program. Once a panel has been tested and analyzed as a 'practice' panel, it cannot be reclassified for use in laboratory qualification. Results from these 'practice' panels will be assessed and results returned to the laboratory, but no proficiency score will be assigned.

### Change in Status Letters

A laboratory will receive a change in status letter if they obtain a score on a round of testing that changes their overall performance rating. This letter will document the laboratory's scores over the last three rounds of testing and will indicate when a change in status (performance rating) has resulted. A copy of this letter will be sent to the director of the laboratory and the network laboratory group for whom the laboratory does testing, as appropriate. Letters will be sent to notify individuals of both negative and positive changes in performance ratings. The VQA submits the letters on behalf of the VQAAB, but has no control over the implementation of rules governing the ability of a site to continue protocol testing. All questions surrounding a laboratory's ability to resume or discontinue protocol testing should be directed to the respective network laboratory group or leadership.

## VQA Program Quantitative HIV-1 RNA Proficiency Testing Scoring Procedure

### Scoring Criteria for VQA Quantitative HIV-1 RNA Proficiency Testing

The VQA has determined that a minimum of 20 samples are needed to provide sufficient power for analyses. With only five samples on the new panels, no single panel could provide information on all of the scoring criteria. For example, a panel of five samples at the same nominal concentration would provide information on reproducibility and validity, provided the nominal concentration was above the range in which assay precision varies inversely with RNA concentration. However, this panel would not provide information on assay specificity or sensitivity. In most cases, statistical power to identify problems would also be quite limited given the small size of each panel. Therefore, proficiency scores will be based on the combined results from four consecutive panels, or 20 total samples. This approach boosts statistical power and allows performance to be monitored over an interval of twelve months rather than at one point in time.

The scoring criteria have been designed to reflect the way that RNA assays are used in clinical trials. Viral load measurements are currently conducted in real-time. Therefore, assessments of the total assay standard deviation, including both the intra-assay and inter-assay components, is used to monitor proficiency. In addition, testing for accuracy has been added to the performance criteria. Finally, sensitivity, specificity, assay validity, data timeliness and data query responsiveness are also used for measuring performance. Methods for assessing performance are discussed below.

Technical Performance

1. Precision

As noted earlier, the measurement of change within a patient is not affected by inter-assay variation when all of the samples from that patient are assayed in a single batch. However, both intra-assay and inter-assay variation are important if samples are assayed in real time. The combination of the two components is defined here as total assay variation. When the original proficiency testing program for RNA assays was designed, the committee of virologists who worked with the VQA Program to develop performance criteria specified that the assays should be conducted with precision sufficient to detect a 5-fold difference between two samples in the same batch. This was interpreted to mean that the intra-assay standard deviation should provide at least 90% power to detect a five-fold difference, which led to the criterion of an intra-assay SD that was not statistically significantly greater than 0.15 $\log_{10}$. A new target standard deviation that takes account of both intra-assay and inter-assay variation is needed for the revised program. However, no mandate analogous to detecting five-fold differences has been provided for assay variation in real time testing. The statistical test for assessing performance also posed a difficulty. Under the original program, the

intra-assay variance, which is the square of the intra-assay SD, was compared with the square of 0.15 $\log_{10}$ using a chi-square statistic. It is not possible to perform a similar test on the total assay standard deviation, because the square of the total assay standard deviation does not follow a chi-square distribution. In fact, it does not have a closed form solution that would lead to simple calculation of a test statistic.

Another problem that arose was the need for replication within each five-member panel if the intra-assay component of variation was to be assessed. Under the original program, the intra-assay standard deviation was estimated from the variation in estimated RNA concentration among coded replicates at the same nominal concentration within a 20-member panel. Replication at even one nominal concentration within a five-member panel was considered too restrictive.

These problems were solved by replacing the original system for assessing assay precision with the following approach. (1) Performance is now assessed using $\log_{10}$ recovery, rather than $\log_{10}$ estimated RNA concentration, where recovery is the ratio of estimated to nominal concentration. This choice was based on the assumption that recovery is constant within the linear range of an assay. Under this assumption, recoveries can be combined across nominal concentrations within five-member panels to provide the replication needed for estimating the intra-assay component of variation. (2) In the absence of a mandate analogous to detecting a five-fold change, a new target standard deviation was developed by estimating expected values for the intra-assay and inter-assay standard deviations using data from recent proficiency panels. The expected values were combined to provide an expected value for the total assay SD. (3) The expected distribution of the total assay standard deviation for each set of four consecutive panels was determined by Monte Carlo simulation that will be based on the expected values. The simulation process is discussed below. Estimated total assay standard deviations that exceed specified cut points on this distribution will be flagged.

Expected values for the intra-assay and inter-assay standard deviation were actually obtained for the Standard Monitor assay using data from the 'A' rounds of panels 016r through 019r. Using data from panels that received passing scores, the intra-assay SD was estimated to by 0.1196 $\log_{10}$and the inter-assay SD was estimated to be 0.0831 $\log_{10}$. A recent update, using data from the 'A' rounds of panels 016r through 022r produced very similar values. Using the 'A' rounds of ultralow proficiency panels 004ru through 009ru, the intra-assay and inter-assay SD's for the Ultrasensitive Monitor assays were estimated to be 0.1196 $\log_{10}$ and 0.0894 $\log_{10}$ respectively. This analysis was also restricted to data sets with passing scores. The inter-assay SD was slightly higher on the Ultrasensitive Monitor assay than on the Standard Monitor assay because of some problems with low recovery in a few laboratories on a few rounds of testing. For example, problems with a centrifuge were identified at one site. Therefore, the values from the Standard Monitor assay were treated as empirically determined expected values for the intra-assay and inter-assay standard deviation. Taken together, they imply that the expected value for the total assay standard deviation is:

$$(0.11962+0.08312)\ 1/2 = 0.1456\ \log_{10}$$

As noted, these estimates were obtained using only the panels that received passing scores. Panels that had minor or major problems were eliminated for the following reasons. The most common problem encountered on proficiency panels is an elevated standard deviation. The elevated values appear as outliers in a plot of the frequency distribution of standard deviations. It did not seem appropriate to base the estimated standard deviations on a data set that included values that did not fit the overall distribution. Furthermore, the residuals from the linear model that were used to obtain the two standard deviations did not fit the assumption of normality when the complete data set was used to obtain the estimates. This failure was caused by extreme values in the data sets; i.e. by outliers that did not fit the overall distribution of HIV RNA results.

The same expected values for the two components of variation will be used for all kits rather than separate expected values for each kit. This is a continuation of past practice.

The Monte Carlo simulations that are used to derive the distribution of the total assay SD from the expected intra-assay and inter-assay SDs proceed as follows. Suppose that the total assay standard deviation was evaluated using three samples from each of four consecutive panels (i.e. each panel also includes a negative sample or a sample near the limit of detection of the assay). There are five steps to the simulation process:

1. A random sample of size 4 is selected from a normal distribution with SD = 0.12 $\log_{10}$. This standard deviation represents expected intra-assay variation; i.e. the variation among results on the same panel.
2. Step 1 is repeated three more times to make four groups of 4 samples each.
3. A random sample of size 4 was selected from a normal distribution with SD = 0.084 $\log_{10}$.
4. The first value in step 3 is added to each of the four values in the first group, the second value in step 3 is added to each of the four values in the second group, and so on. The values generated in step 3 represent the added contribution of run-to-run, or inter-assay, variation to total assay variation; i.e. this component affects all four samples in a given group of 4 the same way but it varies among groups.
5. Estimate the inter-assay and intra-assay standard deviations from the 16 simulated values. Calculate the total assay SD from these estimates.

This process was repeated 10,000 times to generate a distribution for the total assay standard deviation (**Table 3**). The median was very close to the expected value that was defined above. The cut off points at the 95th and 99th percentiles would be used to identify relatively minor and more serious problems with assay precision.

**Table 3: Percentiles of the distribution of 10,000 simulated total assay standard deviations, assuming four proficiency panels with four samples each.**

| MINIMUM | 25th | MEDIAN | 75th | 95th | 97.5th | 99th | 99.5th | MAXIMUM |
|---|---|---|---|---|---|---|---|---|
| 0.050 | 0.122 | 0.142 | 0.164 | 0.199 | 0.213 | 0.226 | 0.241 | 0.358 |

The results of a given set of simulations will depend upon the total number of samples included in an evaluation of reproducibility and the distribution of these values among panels. For example, different cut points would be obtained if the 16 samples were distributed among the four panels such that two panels had five each and the other two had three each.

The total assay standard deviation (SD) is related to the intra-assay and inter-assay standard deviations as $S_T = (S_A^2+S_E^2)^{1/2}$ where $S_T$ is the total assay SD, $S_A$ is the intra-assay SD and $S_E$ is the inter-assay SD. The total assay SD is calculated for each data set generated within a laboratory using a given kit. The calculated targets for the 95$^{th}$ and 99$^{th}$ percentile for each combination of panels are used to determine the precision score for that laboratory/kit/panel combination with scores below the 99$^{th}$ percentile and equal to or above the 95$^{th}$ percentile receiving a score of Satisfactory with a PIA; scores exceeding the 99$^{th}$ percentile receiving a score of Unsatisfactory.

### 2. Accuracy

Accuracy is assessed by comparing average (mean) $\log_{10}$ recovery in a laboratory with expected $\log_{10}$ recovery. The average is calculated using the subset of samples that are used to assess the total assay standard deviation. Median $\log_{10}$ recovery across all laboratories is treated as the expected value. Recovery is used instead of estimated RNA concentration because recoveries can be combined across

nominal concentrations within a panel or across successive panels to assess accuracy. The median across all laboratories will be used as the expected value, instead of the mean or nominal concentration, for the following reasons. Use of a nominal concentration presupposes that expected recovery should be 100%. This is not reasonable if different kits on average produce somewhat different estimates. Means are more sensitive to the effects of outliers than medians and very large outliers have occasionally been identified on previous panels.

The test statistic for identifying accuracy problems is the distance, in standard errors, between average $\log_{10}$ recovery in a laboratory and expected $\log_{10}$ recovery: (R-M)/S where R is average $\log_{10}$ recovery in a laboratory, M is median $\log_{10}$ recovery across all laboratories and S is the standard error of the mean. A normal distribution is assumed for the test statistic. The standard error is derived from the inter-quartile range of the distribution of mean $\log_{10}$ recovery among laboratories; if the mean is normally distributed, then the inter-quartile range is 1.35 standard errors.

Accuracy statistics have been scored officially since September 2011. In some analyses, it has been noted that the standard error that is used for scoring accuracy does vary by round. If the standard error becomes too small, then there is an increased chance that data sets would be penalized for accuracy when, in fact, the deviations noted in the analysis were quite small. Per the VQA Advisory Board's (VQAAB) recommendation, an analysis investigating the impact of standard errors on the accuracy statistics of previous rounds of testing was conducted in April 2009. As a result of this investigation, a modified approach for the scoring accuracy was implemented. Under this modified approach, the observed standard error on a round of testing would be replaced by the historical median of 0.080517 when the observed SE was below the median; otherwise, the observed value would continue to be used

Accuracy statistics that indicate a departure of at least three but less than four standard errors from the expected value will result in a score of Satisfactory with a PIA, and accuracy statistics that indicate a departure of four or more standard errors from the expected value will result in an Unsatisfactory score.

### 3. Sensitivity

Sensitivity is assessed using a sample containing 50 copies/mL and determining the number of false negative results obtained from replicates at the limit of detection for the assay in question. The maximum acceptable number of false negative results is the number of false negatives that would be exceeded no more than 5% of the time, given the number of samples (3–5) involved and assuming independent binomial sampling with an underlying probability of a negative result of 0.05. Two or more false negatives at 50 copies/mL out of four results in an Unsatisfactory score.

### 4. Specificity

Specificity is assessed by monitoring the results obtained on samples that do not contain HIV. Each panel includes a small number of HIV-negative samples. Negative results (HIV RNA not detected) are expected from all of them; qualitatively detected HIV RNA is deemed a false positive result. One or more false positives results in an Unsatisfactory score.

### 5. Assay Validity

It is expected that all results should meet established criteria for assay validity. The criteria vary among kits. If the manufacturer-defined controls fail this will invalidate the run and the run must be repeated prior to data submission. If an assay censors a sample within a valid run, then the sample must be repeated. Only the repeated valid sample will be included in the analysis. Only valid data

should be submitted. Submitted data containing one invalid result per round will receive a score of Satisfactory with a PIA. Two or more invalid results per round will receive an Unsatisfactory score.

6. Data Entry Errors

Data for each Quantitative HIV-1 RNA PT is submitted using the VQA web-system, which requires entry of viral load data and recommended upload of raw data files. Participating sites have the ability to save their data in the web-system and review it before submission. Prior to the PT due date, a site may request to re-open their PT to correct data entry errors. However, sites will not be able to correct any data entry errors after the PT due date. Data that are submitted to the VQA that contains data entry errors such as entering the wrong values, putting decimals in the wrong place, entering data in the wrong location, switching samples, or submitting the wrong file may result in a score of Unsatisfactory.

## Proficiency Scoring Format for VQA Quantitative HIV-1 RNA Proficiency Testing

The scoring process is illustrated by the example in **Table 4**. The five columns in the table represent the nominal concentrations on five consecutive 5-member panels that are designed to assess proficiency on a quantitative HIV RNA assay. The first four panels would be used to generate the first proficiency score. Specificity would be assessed using the negative samples, sensitivity would be assessed using the samples with a nominal value of 50 copies/mL, while and accuracy and reproducibility would be assessed using the samples with nominal concentrations of 100 copies/ml or higher from all four rounds. The next proficiency score would be based on the combined data from rounds 2 through 5 once the results of round 5 were available.

While a single five-member panel can provide only limited information, it could provide important information if performance problems are severe. Performance on individual panels will therefore be monitored. Laboratory personnel and the VQAAB will be kept apprised of results.

**Table 4: Five hypothetical 5-member panels (Entries are nominal concentrations).**

| Round 1 | Round 2 | Round 3 | Round 4 | Round 5 |
|---|---|---|---|---|
| 0 | 300,000 | 3,000 | 15,000 | 0 |
| 15,000 | 0 | 50 | 3,000 | 10,000 |
| 50 | 15,000 | 300,000 | 0 | 50 |
| 3,000 | 30,000 | 0 | 50 | 400,000 |
| 300,000 | 50 | 30,000 | 30,000 | 60,000 |

### Determination of Proficiency Scores

A summary of the VQA Quantitative HIV-1 RNA PT Scoring is found in **Table 5**. Briefly, a score of Satisfactory will be assigned if the run is valid and no problems with the data were detected.

If minor problems are identified a laboratory will still be assigned a score of Satisfactory, however, they will also receive a Potential Issue Alert (PIA). PIA will be released if:

- SD between 95$^{th}$ – 99$^{th}$ percentile (Monte Carlo Simulations per round);
- Accuracy score indicates $\log_{10}$ recovery differs by at least 3 but less than 4 standard error;
- One Invalid result per round

A score of Unsatisfactory will be assigned for the following criteria:

- a total assay SD that exceeds the 99th percentile on the Monte Carlo simulations;

- an accuracy score that indicates that average $\log_{10}$ recovery differs from expected $\log_{10}$ recovery by at least 4 standard errors per round;
- one or more false positives per round;
- two or more false negatives at the Limit of Detection (50 copies/mL);
- one or more false negatives with samples >50 copies/mL;
- data submitted late without having received prior approval to be exempted from the PT round

**Table 5: Summary of VQA Quantitative HIV-1 RNA PT Scoring**

| RNA Criteria | Score |
|---|---|
| • Run was valid<br>• No problems were detected | Satisfactory (S) |
| • SD between 95th – 99th percentile (Monte Carlo Simulations per round) | Satisfactory (S) +PIA |
| • Accuracy score indicates $\log_{10}$ recovery differs by at least 3 but less than 4 standard error | Satisfactory (S) +PIA |
| • One Invalid result per round | Satisfactory (S) +PIA |
| • One or more False Positives | Unsatisfactory (U) |
| • Two or more False Negatives at Limit of Detection (50 copies/mL) | Unsatisfactory (U) |
| • One or more False Negatives with samples at >50 copies/mL | Unsatisfactory (U) |
| • Two or more Invalid Results per round | Unsatisfactory (U) |
| • SD exceeds 99th percentile (Monte Carlo Simulations per round) | Unsatisfactory (U) |
| • Accuracy score indicates $\log_{10}$ recovery differs by at least 4 standard error | Unsatisfactory (U) |
| • Data not submitted or submitted after the PT due date | Unsatisfactory (U) |

If a score of Unsatisfactory is assigned and the problem can be traced to the most recent 5-member panel, then a re-coded version of that panel will be sent to the laboratory. Results of the repeat assays will be used in all subsequent scoring. Suppose, for example, that an Unsatisfactory is assigned because of problems on the fourth of the five hypothetical panels in **Table 4**. The score for the first four consecutive panels would be based on the first run of round 4. The score for rounds 2 through 5 would be based on the second run of round 4 and the first runs of rounds 2, 3, and 5. If a laboratory receives a Potential Issue Alert caused by the most recent panel, then the panel will be repeated at the discretion of the Laboratory Director.

If the problem exists over multiple panels and repeating the last 5-member panel will not resolve the problem, then the laboratory will be given the option of completing a requalification panel (20-member panel). This will permit the laboratory to replace all previous data for future analyses. As

with a new laboratory running a prequalification panel, as new 5-member panel data are added to the analysis, 5 samples from the requalification panel will be removed in sequential order until only 5-member panel data are used in the analysis.

*Note*: *only intra-assay precision statistics will be used when analyzing a 20-member panel. Total assay standard deviations will be implemented once the first 5-member panel has been added to the analysis.*

After the analysis is complete, and the scores have been approved by the VQAAB (if needed), a report will be uploaded in the VQA web-system. This summary will include the decoded data from that laboratory as received by the VQA, performance on each scoring criterion, and an approved score. The laboratory must contact vqa-rna@duke.edu via e-mail if there is a discrepancy in the report received from the VQA.