Assay Migration Studies for In Vitro Diagnostic Devices

Guidance for Industry and FDA Staff

Document issued on: April 25, 2013

The draft of this document was issued on January 5, 2009.

For questions regarding this document contact Sally Hojvat, PhD, CDRH/OIR/DMD, by phone at (301) 796-5455, or by email at sally.hojvat@fda.hhs.gov or Marina V. Kondratovich, PhD, CDRH/OIR by phone at (301) 796-6036 or by email at marina.kondratovich@fda.hhs.gov. For questions relating to devices regulated by CBER, contact the Office of Communications, Outreach and Development, CBER at 301-827-1800 or 800-835-4709.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of In Vitro Diagnostics and Radiological Health
Center for Biologics Evaluation and Research

Preface

Public Comment

You may submit written comments and suggestions at any time for Agency consideration to the Division of Dockets Management, Food and Drug Administration, 5630 Fishers Lane, rm. 1061, (HFA-305), Rockville, MD, 20852. Submit electronic comments to http://www.regulations.gov. Identify all comments with the docket number FDA-2008-N-0642. Comments may not be acted upon by the Agency until the document is next revised or updated.

Additional Copies

Additional copies are available from the Internet. You may also send an email request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance, or send a FAX request to 301-847-8149 to receive a hard copy. Please use the document number 1660 to identify the guidance you are requesting.

Additional copies of this guidance document are also available from the Center for Biologics Evaluation and Research (CBER) by written request, Office of Communication, Outreach and Development (HFM-40), 1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448, by telephone, 1-800-835-4709 or 301-827-1800, by email, ocod@fda.hhs.gov, or from the Internet at http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/General/default.htm.

Table of Contents

Assay Migration Studies for In Vitro Diagnostic Devices

Guidance for Industry and FDA Staff

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. Introduction

This guidance outlines FDA's regulatory approach for approval of Class III in vitro diagnostic devices, as well as certain licensed[1] or certain cleared[2] in vitro diagnostic devices in cases, when a previously approved, licensed, or cleared assay is migrating (i.e., transitioning) to another system for which assay performance has not been evaluated by FDA.[2] In this guidance the term "new system" refers to the system (assay, instrument, and software) to which the assay is migrating from a previously approved/licensed/cleared system. The term "old system" refers to the system (assay, instrument and software) which was approved/licensed/cleared from which the assay is migrating to a currently unapproved/unlicensed/uncleared system.

This guidance is focused on the study designs and performance criteria that should be fulfilled in order for a sponsor to utilize the migration study approach in support of the change. FDA will review information from the sponsor, including results of analytical and comparison studies outlined in this guidance, as well as device descriptions and risk analyses, to determine whether the use of the approved/licensed/cleared assay with the new system may compromise safety and effectiveness of the assay. The guidance document describes information that we recommend you include in a PMA (premarket approval

[1] This guidance does not apply to immunohematology tests licensed by the Center for Biologics Evaluation and Research (CBER).

[2] This guidance can be used for 510(k) devices where the Replacement Reagent and Instrument Family Policy (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079185.htm) does not apply (e.g., nucleic acid amplification tests) and devices for which transitioning to a new system presents specific concerns, either because of the nature of the analyte and indications, or because of the specific technology used.

application) supplement, a BLA (Biologics License Application) supplement, or a 510(k) (premarket notification). For devices regulated by OIR (Office of In Vitro Diagnostics and Radiological Health, within CDRH), sponsors may contact OIR, and for those regulated by CBER, sponsors may contact CBER to obtain feedback concerning study plans.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. Background

The FDA believes that the assay migration study paradigm discussed in this guidance provides a least burdensome scientific and regulatory pathway for manufacturers to transfer a previously approved or licensed assay with full clinical data from an old system to a new system (not approved. licensed, or cleared). The paradigm is suitable in cases when sufficient knowledge can be derived from the documentation of design controls, risk analyses, and prior performance studies on an old system.

If you make further modifications or iterations of the old or new system you should compare back to the original old system that has full clinical data when performing new migration studies. However, if the old system with full clinical data is no longer available please contact the appropriate FDA CDRH or CBER Division for further discussion.

The migration studies approach is related to the Replacement Reagent and Instrument Family Policy that FDA uses for many Class I and Class II diagnostic devices.[3] Similar to that policy, migration studies rest on the premise that as platform changes are made throughout the lifetime of an approved, licensed, or cleared assay, smaller and more focused analytical and clinical data sets, along with prior knowledge of device design and performance, could allow for a determination that the safety and effectiveness profile of the modified system has not been compromised.

The assay migration study paradigm is practical, risk based, and consistent with FDA's Critical Path Initiative, which is intended to help bring new medical products to market successfully and efficiently[4]. FDA believes that with proper control and review, migration studies will meet regulatory thresholds for premarket review in a manner that will be least burdensome for both companies and FDA while protecting public health.

[3] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079185.htm.
[4] http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html

III. Scope

This guidance is intended to be applied, where appropriate, to licensed donor screening tests[5] and approved (Class III) in vitro diagnostic assays, as well as 510(k) cleared assays for which migration to a new system presents concerns. Possible scenarios for assay migration are when the assay is being transferred from a manual system to an automated or semi-automated instrument system or from a semi-automated instrument system to a fully automated instrument system, or from one automated instrument system to another (and vice versa for all scenarios). A broad variety of methodologies may use the migration studies paradigm depending on what is known about the two Systems involved. Assay transfer may be from one approved, licensed, or cleared old system to a new system that has the same technical characteristics. However, if scientific evidence suggests migration studies may not be adequate to predict actual clinical testing performance on the new system, the assay migration paradigm should not apply and FDA will recommend that traditional evaluation studies be performed.

Assay migration studies are ideally suited for test systems for which the assay output (raw signal) is a numeric result or is expressed as a signal to cutoff (S/CO). Devices for which a numeric output is not available might be more difficult to analyze and may not be suitable candidates for use of this approach.

Migration studies would not be applicable to the following devices or to system changes that are generally considered significant, such as:

- systems intended for over-the-counter use,
- systems intended for prescription home use, or
- devices that do not meet the Critical Considerations criteria stated below.

The FDA strongly recommends that sponsors discuss proposed migration studies with the Agency early in the product design and testing process in order to determine if the proposed changes are consistent with the considerations that would allow for focused testing. An early contact with the Agency is especially recommended for an assay being transferred from a manual system to an automated or semi-automated instrument system. For CBER licensed devices, this contact may be through an IND, or protocol review, providing preliminary protocols, data, and justifications prior to performing the migration studies. The size, nature,

[5] FDA does not believe that this guidance is suitable for use in its entirety when immunohematology tests (e.g., blood grouping, blood group antibody detection and/or identification, crossmatching) are being migrated because of the differences in assay methodology and results reading and interpretation as compared to the other assays and systems described in this guidance. If you believe that your immunohematology reagents and system can be evaluated using the criteria outlined in this guidance, contact the responsible review division in CBER. Immunohematology products are reviewed in the Devices Review Branch, Division of Blood Applications, Office of Blood Research and Review, CBER.

and scope of migration studies we recommend will depend on a detailed evaluation of the old and new systems, the level of regulation appropriate for the product (e.g., PMA, BLA, or 510(k)), the performance characteristics of the assay, and the design and scope of the analytical testing and clinical trials used to support approval, licensure, or clearance of the assay on the old system.

IV. Critical Considerations for Determining Whether The Migration Studies Paradigm May Apply To A Particular Device

A sponsor should take into account the following critical considerations in determining whether the migration studies approach is appropriate for a particular product, and describe these considerations in the submission:

- The intended use and indications for use for the new system should be unchanged from the old system, except for inclusion of the new system.

- Reagent and assay parameters (e.g., cutoff) should be unchanged, except for very minor differences (such as small changes in incubation times) in order to optimize the assay on the new system. However, the sponsor should provide evidence that the changes do not compromise the assay's performance.

- Some assay technologies may not be good candidates for the migration studies paradigm (e.g., assays with relatively high imprecision near the assay cutoff).

- Assay and system technologies should remain unchanged. All biochemical (e.g., antibody and antigen interactions or DNA probe construct) and physical detection (e.g., colorimetric, chemiluminescence, or dye binding) technologies should be unchanged from the old system. Minor differences in hardware instrumentation may be appropriate and will be evaluated on a case-by-case basis.

- There should be no expected change to the assay performance when run on the new system. However, actual changes will be evaluated in the context of their impact on the clinical use of the assay. Due to the limited number of positive and negative samples, the migration studies approach is not appropriate to support changes in clinical performance claims.

V. Additional Considerations

In addition to addressing each of the critical considerations noted in Section IV, you should also include the information listed below to demonstrate the applicability of migration studies to the transfer of the assay from the old system to the new system. The information in your submission should include, but is not limited to, the following:

- Device description, including functional block diagrams and hardware and software components for both the old and new systems to allow an evaluation of the changes to the new system when compared to the old system.

- A similarities and differences table for a side-by-side comparison of assay parameters and hardware/software functions, requirements and design.

- A risk analysis of the new system (software/hardware/assay) using relevant guidance documents.[6,7]

- Summaries of software data validating functional operation of alerts and alarms in real or simulated circumstances.

- System Operator Manual(s): If the new system was previously approved or licensed with analytes other than the one under consideration, a new review of the System Operator Manual may only be called for if changes have been made that impact safety and effectiveness, or if there are assay-specific sections or changes.

- Proposed modifications to the labeling that appropriately describe respective prior data and new migration study information. The sponsor should consult with FDA when determining whether a dual or separate package insert will be appropriate. Inclusion of the old system's analytical and performance data should be included where appropriate.

In addition, the sponsor should include documentation on software and instrumentation for the new system. When appropriate for the device, this documentation should meet all recommendations for the appropriate Level of Concern.[7] To further minimize risk, studies should be performed on the final production model of the new system that is intended to be marketed. Likewise, studies should be performed using the final release candidate of the software and any differences between the release candidate and the final release should be evaluated using a risk based approach and included in the submission for FDA review.

If you are using the migration paradigm to migrate an approved Class III assay to an instrument that has not previously been cleared or approved, or to move an approved Class III manual assay to an instrument not previously cleared or approved, you should submit documentation describing the hardware and software design controls, including design verification and validation, along with other appropriate Quality Systems documentation. (See 21 CFR Part 820 – Quality System Regulation). For a general discussion of the Quality System information that FDA recommends including in premarket applications please also refer to the guidance document available at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/uc

[6] ISO 14971:2007, Medical devices – Application of risk management to medical devices

[7] Guidance for Industry and FDA Staff: Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm.

m070897.htm. In addition, if the instrument manufacturing site has not been inspected as part of the original class III assay approval process, then a site inspection may be conducted.

VI. Assay Migration Studies

This section outlines specific studies that may be appropriate to support assay migration for in vitro diagnostic devices. Before preparing to use the migration studies approach, you should determine whether the assay is quantitative, qualitative or semi-quantitative, according to the definitions in this guidance. Specifically, for the purposes of this guidance, qualitative assays are those that use numeric values (e.g., signal, S/CO), to determine nominal categorical assay results (e.g., positive or negative). Quantitative assays determine numeric values which are referenced to a measuring interval, and standards that allow determination of analyte concentrations. Section VI.A below describes studies for qualitative assays; Sections VI.B and VI.C address quantitative assays and semi-quantitative assays respectively; and Section VI.D addresses point-of-care assays. Special considerations for blood screening assays are covered in Appendix I, "Migration Studies for Blood Donor Screening Assays."

A. Migration Studies for Qualitative Assays

1. Analytical Studies for Qualitative Assays

The evaluations described below are based on the idea that similar studies were conducted previously for the old system. If the study design of the analytical studies conducted for the old system were different from the design of the studies described in this guidance, please contact the FDA for feedback. If you believe that some of the studies outlined in this document do not apply to your particular device, you should present your justification for FDA review.

We recommend that you use fresh clinical specimens for all analytical studies. If this is impractical, in some cases you may substitute or supplement fresh clinical specimens with archived samples. If archived samples are not available, spiked or diluted clinical samples may be used. In some instances, use of otherwise contrived matrix-specific samples may also be appropriate; however these should mimic clinical specimens as much as is feasible. We recommend that you contact FDA if you wish to discuss appropriate sample types for these evaluations. The matrix of any of these alternative specimens should be the same as that specified by the intended use of the old system.

a) Performance at Low Analyte Levels

You should evaluate the performance of the assay on the new system compared to the old system at low analyte levels (e.g., with dilution panels and seroconversion panels).

- Where available, assay performance at low positive analyte levels using

dilution panels should be based on international standards (e.g., World Health Organization (WHO) standards, Paul Ehrlich Institute (PEI) standards) and compared with the old system.

- If available and appropriate, test well-characterized seroconversion panels similar in number and type to the panels originally used to support approval/licensure/clearance. The seroconversion panels should be run on both the old system and the new system.

b) Precision

This section describes precision study designs including panel composition.

(i) Composition of Precision Panel

The categorical result of a qualitative test is determined using the numeric test data and a cutoff(s) (e.g., a numeric threshold), which define numeric ranges that correspond to the final result categories. For a qualitative test with two categorical results, a cutoff (one threshold) is used to define positive and negative results of the test. When the observed numeric test data exceeds the threshold, the final result is considered positive (or "reactive", "detected"). When the observed numeric test result is below the cutoff, it is considered negative (or "not reactive", "not detected").

Qualitative assays discriminating between two classes of subjects (non-diseased vs diseased) can fall into two groups: 1) assays used in situations where non-diseased subject samples have a true zero concentration of the analyte of interest, and 2) assays used in situations where non-diseased subject samples have a detectable analyte concentration. For the latter, the cutoff was chosen to optimize clinical sensitivity and clinical specificity of the assay based on a clinical data set. Appropriate precision panels for these two types of assays, as well as assays that produce equivocal results, and/or have a re-test zone, are discussed in more detail in ***(a)-(c)***, *below*:

(a) When non-diseased subject samples have a detectable amount of analyte.

For the non-diseased subject samples that have a detectable amount of analyte the cutoff of the assay is higher than the limit of the blank. A useful characteristic of the cutoff is that a sample with a concentration (determined by the mean signal to cutoff value from a large a series of replicate measurements) at the cutoff yields a positive result 50% of the time and a negative result 50% of the time when a large number of replicates of that sample are run under stipulated conditions (see Figure 1 below). We denote this concentration as C_{50}.

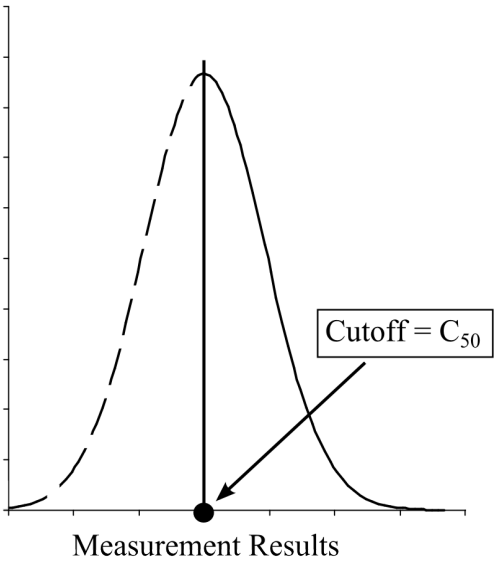

Figure 1. Results of a qualitative assay (probability density function) for a sample with a concentration at the cutoff measured multiple times.

For samples with concentrations exceeding C_{50}, one expects to see positive results more than 50% of the time, and similarly for samples with a concentration below C_{50}, one expects to see positive results less than 50% of the time. In this guidance we refer to an analyte concentration that yields, upon evaluating many replicates, a positive result 95% of the time (and a negative result 5% of the time) as a Low Positive concentration (C_{95} concentration). We refer to a sample concentration which yields a positive result 5% of the time (and negative result 95% of the time) as a High Negative concentration (C_5 concentration). Samples with concentrations of analyte close to C_{95} and C_5 as determined by the old system are recommended for the within-laboratory precision (see CLSI[8] document EP12). The panel should consist of at least four members, as described below (also, see Figure 2):

- A True Negative sample: a sample with a true analyte concentration of zero.

- A High Negative sample: a sample that repeatedly tests negative approximately 95% of the time and positive 5% of the time by the old system. One should expect that the same concentration tested by the new system will also produce negative results approximately 95% of

[8] Clinical Laboratory and Standards Institute. CLSI documents referenced in this guidance are listed in Section XI of this document.

the time (C_5 concentration as determined by the old system).

- A Low Positive sample: a sample with an analyte concentration around C_{95} as determined by the old system. Repeatedly testing this sample by the old system should give a positive result approximately 95% of the time and a negative result 5% of the time. One should expect that the same concentration tested by the new system will also produce positive results approximately 95% of the time. Note that if the LoB is used as a cutoff, then the concentration C_{95} is the same as the LoD.

- A Moderate Positive sample: a sample with a concentration close to the cutoff, and at which one observes positive results by the old system approximately 100% of the time (e.g., a sample with a signal approximately two to three times the signal at the cutoff if the cutoff=1.0, or a sample with concentration approximately two to three times the 95% LoD if the cutoff is based on LoB).

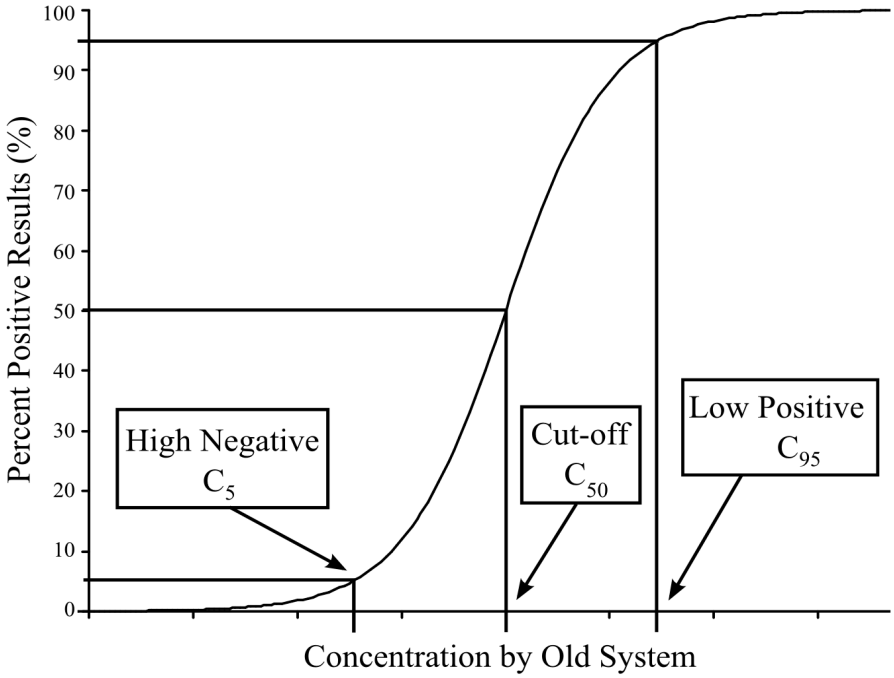

Figure 2. Relationship between percent of positive results and the analyte cutoff concentration.

For the details of how the C_{95} and C_5 concentrations can be evaluated from the previous precision studies of the old system, see Appendix II: Statistical Notes, 1. For the precision study of the new system, it is not necessary to have the High Negative and Low Positive samples at exactly C_5 or C_{95} of the old system. If the High Negative and Low Positive samples in the precision study of the new system are close enough to the cutoff that the standard

deviation (or percent coefficient of variation (CV) is approximately constant over the range around the cutoff, the C_5 and C_{95} of the new system can be evaluated from this precision study (see Appendix II: Statistical Notes, 1).

(b) When non-diseased subject samples have a true zero analyte concentration.

When non-diseased subject samples have a true zero analyte concentration there are two types of assays: (1) assays for which samples with zero concentration have a distribution of numeric signals and the cutoff of the assay is the limit of blank (with some type I error, for example, 5%); and (2) ultrasensitive assays for which samples with zero analyte concentration almost always have a negative assay result (type I error is close to zero). These two types of assays are discussed further in the two bullets below.

- Consider assays where samples with zero analyte concentration have a distribution of numeric signals and the limit of blank is used as a cutoff. In this case the concentration C_{95} is the same as the limit of detection (LoD) and zero concentration (no analyte present in sample) is C_5 as illustrated in Figure 3, below. The precision panel should consist of at least the following three members: True Negative, Low Positive (around LoD) and Moderately Positive.

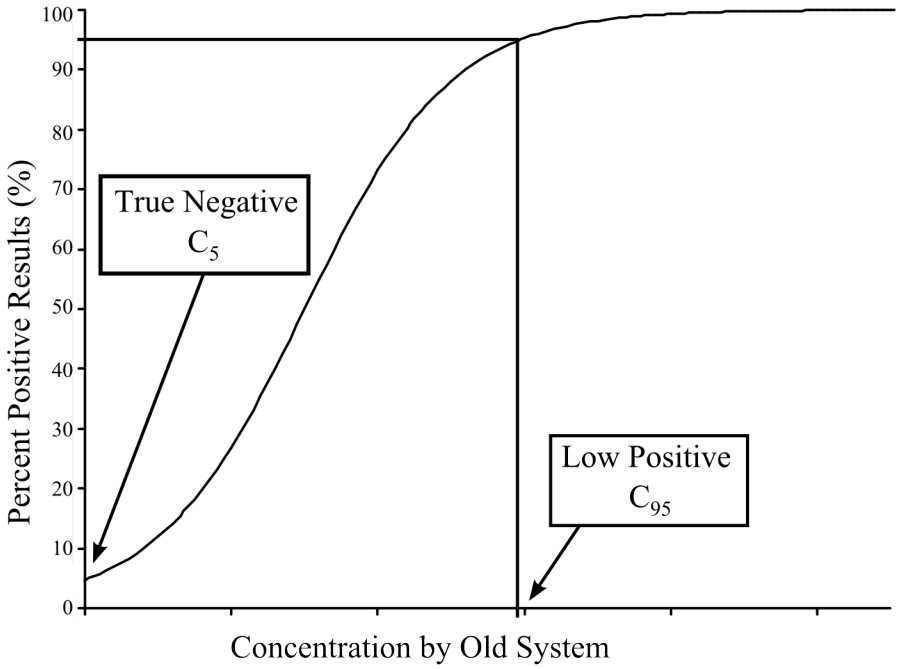

Figure 3. Relationship between the percent of positive results and the analyte concentration where True Negative samples produce numeric results and the C_{95} concentration is the LoD of the assay.

- For ultrasensitive assays that employ, for example, a signal amplification methodology (e.g., real-time PCR), the following approaches for precision panel members are recommended. For ultrasensitive assays, the concentration range that spans High Negative (around C_5) to Low Positive (around C_{95}=LoD) is narrow with respect to the observed concentration range of clinical specimens. Due to this narrow range, it may be difficult to prepare a sample with C_5 concentration.

  If less than 10% of all subjects positive by the old system have an observed concentration less than C_{95}, then samples at the C_5 concentration could be excluded from the precision panel. In this case, samples which compose the precision panel should include the following concentrations: True Negative, Low Positive (around C_{95}) and Moderately Positive.

  If more than 10% of all subjects positive by the old system have an observed concentration less than C_{95}, then samples which compose the precision panel should include the following concentrations: True Negative, Low Positive (around C_{95}), samples in the range of C_{20} to C_{80} and Moderately Positive as illustrated in Figure 4, below.

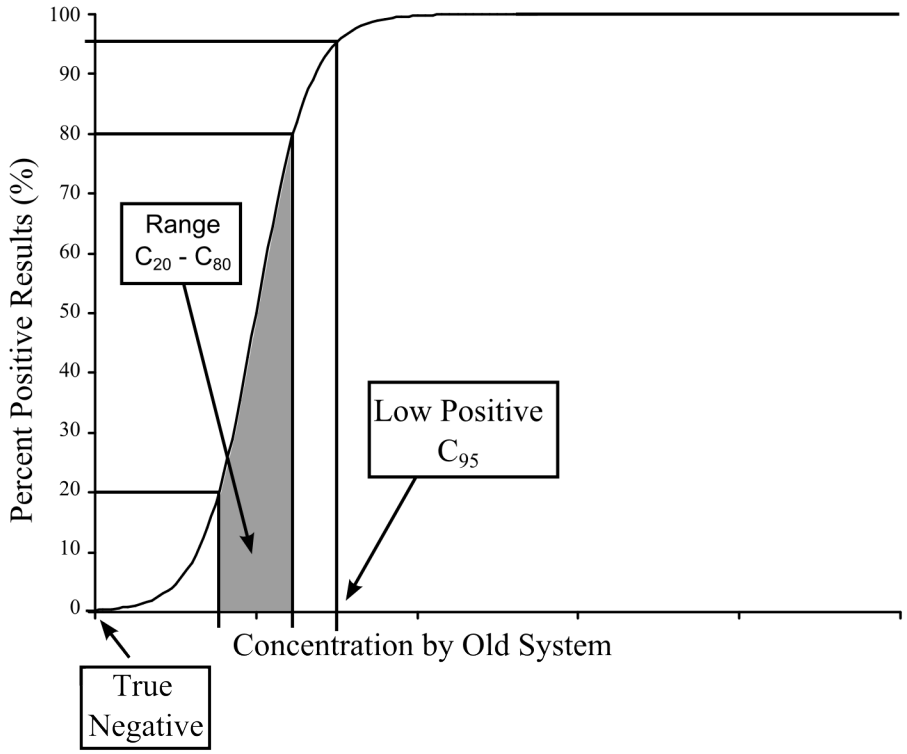

Figure 4. Relationship between the percent of positive results and the analyte concentration for an ultrasensitive assay showing a narrow concentration range between C_5 and C_{95}.

(c) When an assay with a numeric value produces an equivocal result and/or has a re-test zone.

If your assay with numeric values produces an equivocal result and/or has a re-test zone, then the assay has two cutoffs, E_1 and E_2:

If S/CO < E1 then the assay result is Negative

If $E_1 \leq S/CO \leq E_2$ then the assay result is equivocal or in the re-test zone

If $S/CO > E_2$ then the assay result is Positive

As an example, the numeric values of the assay with equivocal results defined as having S/CO from E_1 to E_2 are shown graphically in Figure 5, below.

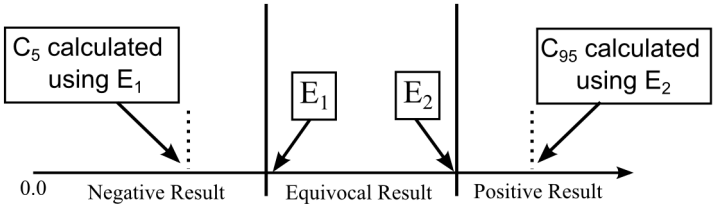

Figure 5. Relationship of numeric values and result categories for an assay with equivocal results.

The C_5 and C_{95} concentrations for the old system are calculated for cutoffs E_1 and E_2 respectively using the same method as described in Appendix II: Statistical Notes, 1. In this case, samples which compose the precision panel should include the following concentrations: True Negative; near the equivocal (and/or re-test) zone: near the C_5 for cutoff E_1 and near the C_{95} for cutoff E_2; and Moderately Positive.

In addition to the samples described for all panels (a-c) above, you should run the appropriate control material (negative and positive controls) and calibrators in the precision study.

(ii) Within-Laboratory Precision Study

We recommend that you conduct in-house within-laboratory precision studies (to supplement the external site reproducibility studies described below in Section VI.A.1.b.iii). When appropriate and justified, the in-house within-laboratory precision study may not be necessary, such as (1) if you established that the new system only needs to be recalibrated at relatively long time intervals (e.g., 6 months or more) and any other concerns can be appropriately addressed by the reproducibility study, or (2) if the new system is recalibrated daily, so that

calibration cycle variability is inseparable from day-to-day variability (which is assessed by the reproducibility studies described below) and any other concerns can be appropriately addressed by the reproducibility study.

It may be sufficient to perform within-laboratory precision studies only on the new system. However, if the study design or composition of the precision panel of the old system precision study was very different from that described in this guidance, it may be important to perform the precision study on the old system as well. The within-laboratory precision study described below is based on CLSI document EP05.

Sources of variability we recommend for the within-laboratory precision study are at least 12 days of testing, with 2 runs per day, and 2 replicates of each sample per run. These 12 days are not necessarily consecutive. If the system calibration variation composes a significant portion of the total system variation, or the calibration interval fits within the time frame of the studies, then at least 2 calibration cycles should be included in the study design; otherwise you should, provide your rationale for not including an evaluation of the variation introduced due to instrument calibration. For each cycle, days at the beginning and end of the cycle should be included (e.g., 3 days at the beginning and 3 days at the end for each cycle). You should include other additional sources of variability in the design of the study, if they are important to the specific assay (e.g., operators, lots, etc.). In such cases, overall modification to the variables might be possible (e.g., spreading days of testing between different operators). If analytical and clinical performance is similar across all matrices that are included in the intended use of the old system, then establishing performance of the new system using the most commonly employed matrix may suffice.

(iii) Reproducibility Study

We recommend you perform the reproducibility study based on a modification of CLSI document EP05 on the new system. The panel composition and analyte levels for this study should be the same as described above (in Section VI.A.1.b.i). Sources of variability should include testing for at least 5 days, 2 runs per day, with 3 replicates of each panel member per run, at 3 laboratories of which at least 2 are external sites. Other sources of variability might be applicable if relevant to the specific assay (e.g. operators). If analytical and clinical performance is similar across all matrices intended for use with the old system, then establishing performance of the new system using the most commonly employed matrix may suffice.

For each concentration level, similar information should be available for the old system. Otherwise, you should perform a new reproducibility study for the old system with study design and concentration levels as described in this section.

2. Comparison Studies for Qualitative Assays

You should perform comparison studies using comparison panels described below. However, minor changes to the old system might not warrant performing all comparison studies. The extent of the utility of these studies can be evaluated on a case-by-case basis in consultation with the FDA.

a) Comparison Panels

For each analyte, the qualitative assay comparison panels should consist of the following:

- Panels of samples known to be positive or negative for a specific assay on the old system (in order to evaluate the same assay on the new system). We suggest that the positive and negative panels have at least 100 panel members each.

  Positive panel members should be prepared so that approximately 60%-80% of the samples have analyte levels close to the cutoff. Of these samples, approximately one half should be close to the C_{95} (or LoD) of the old system and the other half should be Moderate Positive samples. The remaining positive samples should evenly cover the full detection range of the assay.

  Negative panel members should be prepared so that approximately 30%-40% are High Negatives (close to C_5 of the old system). Because FDA is assessing clinical effectiveness based upon the result agreement for the assay performed on the two instruments, these samples can have an analyte concentration below the cutoff such that repeat testing of these samples should be negative approximately 95% of the time. Alternatively, these High Negative samples may be obtained from archived clinical samples giving a signal just below the cutoff. For ultrasensitive assays, negative panel members should include at least 100 True Negative samples.

- If the assay produces an equivocal result and/or has a re-test zone, then the assay has two cutoffs, E_1 and E_2. For the purpose of defining acceptable concentration ranges of negative and positive panel members, consider the lower bound of the equivocal or re-test zone as E_1 and the upper bound as E_2 noted below:

  If $S/CO < E_1$ then the assay result is Negative

  If $E_1 \leq S/CO \leq E_2$ then the assay result is equivocal, or in the re-test zone.

  If $S/CO > E_2$ then the assay result is Positive

Positive panel members should be prepared so that approximately 60%-80% of the samples have analyte levels close to the cutoffs. Of these samples, approximately one half should be close to the $C_{95,}$ calculated using E_2 based on the precision of the old system, and the other half should be Moderate Positive samples. The remaining positive samples should evenly cover the full detection range of the assay. Among the samples close to the C_{95}, some samples should be in the equivocal zone (and/or in the re-test zone), as defined by the old system.

Negative panel members should be prepared so that approximately 30%-40% are High Negatives (close to the C_5 calculated using E_1 based on the precision of the old system). Because FDA is assessing clinical effectiveness based upon the result agreement for the assay performed with the two systems, these samples can have an analyte concentration below the cutoff such that repeat testing of these samples should be negative approximately 95% of the time. Alternatively, these High Negative samples may be obtained from archived clinical samples giving a signal just below the lower bound of the equivocal or re-test zone (E_1). Among the samples close to the C_5, some samples should be in the equivocal (and/or re-test) zone as defined by the old system.

- It is preferable to use non-contrived clinical samples to create the panel members. However, where positive samples are not available or volumes are low, you should work with the FDA to define acceptable sample types. Positive panel members may be diluted in a clinical matrix in some cases. Preferably, individual negative samples should be used for each dilution. Spiked samples may be acceptable, but only from individual strong positives in the appropriate clinical matrix (i.e., the same sample cannot be used to prepare multiple panel samples).

- Archived samples are acceptable if stored in accordance with the package insert instructions. The samples should be randomized and masked.

b) Comparison Study Design

You should test the comparison panels on the old system at a minimum of one site. This may be done in-house. However, we recommend use of at least two of the old systems to better assess systematic differences between different old systems. The new system should be tested at a minimum of three sites (one may be in-house) with one or more reagent kit lots. Each panel member should be tested at least four times: once with the old system and three times with the new system. You should send the same positive and negative panels to each site, rather than dividing the panel between the three sites. Three different builds with unique serial numbers of the new system should be tested, one unique serial numbered system at each of the three sites.

3. Statistical Analysis of Data

We recommend you include the information described in Sections *a-c* below for each of the respective studies, in your submission:

a) Within-Laboratory Precision Study

- For each sample tested, we recommend you present the mean value with variance components (standard deviation and percent CV) for numeric values for the new system. In addition, you should include the percent of values above and below the cutoff for each sample. You should also provide an estimation of C_5 and C_{95} of the new system (see Appendix II: Statistical Notes, 1).

- For each analyte concentration, you should provide similar information for the old system. You can obtain this information from the precision study you originally used to demonstrate the performance of the old system, if the study design and the analyte concentration levels are comparable to those used to assess the new system. You can utilize information about variance components (standard deviation and percent CV) from the precision study found in the old system labeling. However, if the study design and analyte concentrations for the old system are different from those described in this document, a new precision study for the old system should be performed to allow for comparison between systems.

- For similar analyte concentrations, we recommend you provide repeatability (within-run precision) and within-laboratory precision (standard deviation and percent CV) of the New and old systems (precision profiles of the New and old systems). In addition, you should provide the ratio of the standard deviations of the New and old systems along with the 95% confidence interval for this ratio. The confidence interval can be based on the F-statistic for a ratio of variances (see Appendix II: Statistical Notes, 2).

b) Reproducibility Study

- For each analyte concentration level, we recommend you present the mean value with variance components (standard deviation and percent CV) for the new system, for each site separately and for all sites combined. In addition, for each panel member, you should include the percent of values above and below the cutoff, for each site separately and for all sites combined. You should also provide an estimation of C_5 and C_{95} of the new system, for each site separately and for all sites combined.

- For each analyte concentration level, we recommend you present similar information for the old system. This information can be taken from the precision study originally used to demonstrate the performance of the old system if the study design and the analyte concentration levels are comparable to those used to assess the new system. The information about variance components (standard deviation and percent CV) can be obtained from the

precision study found in the old system labeling. If the study design and analyte concentrations are different than those described in this document, a new precision study for the old system should be performed.

- For similar concentrations, you should provide repeatability (within-run precision), within-laboratory precision (standard deviation and percent CV) and reproducibility separately and combined over three sites for the new systems and old system (precision profiles of the new system and old system). In addition, we recommend you provide the ratio of repeatability standard deviations along with 95% confidence intervals and the ratio of the reproducibility standard deviations of the New and old systems. The confidence intervals for the ratios can be based on the F-statistic for a ratio of variances (see Appendix II: Statistical Notes, 2).

c) Comparison Panels

- You should provide a scatter-plot of S/CO or any other numeric values that differentiate between positive (reactive) and negative (non-reactive) of the old system (X-axis) vs. S/CO of the new system (Y-axis) with descriptive statistics, for data generated at each site as well as for all sites combined. Both axes should have the same scale, and the line of identity (y=x) should be presented. The same scale on the axes should be applied to the data from each site. In addition you should provide similar scatter plots for negative and positive specimens separately.

- You should provide an estimation of systematic differences between the numeric values generated by the old system and the new system at each site and for all sites combined (CLSI document EP09). You should perform the appropriate regression analysis (Deming regression, see Appendix II: Statistical Notes, 3), which accounts for the random errors associated with the old and new system measurements and provide the 95% confidence intervals of the slope and intercept from the regression analysis. The emphasis should be placed on estimating the systematic difference between the numeric values of the old and new systems around the cutoff. You should also calculate the average systematic difference separately for the Negative, High Negative, Low Positive and High Positive samples of the comparison panel.

- We recommend that you present tables (e.g., with old system results defined by the columns and new system results defined by the rows), for data derived from each site and for all data combined (see Appendix II: Statistical Notes, 4). For the qualitative assays with equivocal (and/or re-test zone), the results of the comparison study should be presented in 3-by-3 tables from data derived from each site and for all data combined (see Appendix II: Statistical Notes, 5). Based on these tables, you should calculate positive and negative percent agreements at each site along with the corresponding 95% two-sided confidence intervals (for confidence intervals, see Appendix II: Statistical Notes, 6). In addition, you should provide the positive and negative percent

agreements averaged over three sites with the corresponding 95% two-sided confidence intervals. Because the same samples are used at all three sites, we suggest that a bootstrap approach may be useful for calculating such confidence intervals.

- For qualitative assays with equivocal results (and/or a re-test zone), the comparison study data should be analyzed as having three ordinal categories (e.g., negative, equivocal, positive). Note that the purpose of a migration comparison study is to evaluate the extent of differences between the numeric values obtained from the old system and numeric values obtained from the new system. While re-testing may be indicated in the package insert, this additional testing is unnecessary to fulfill the migration comparison study objective. The initial numeric values should be used in the numeric analysis.

  For qualitative assays with equivocal results (and/or re-test zone), we recommend that you present tables of agreement including positive, negative and equivocal (and/or re-test) as categorical results for data derived for each site and for all sites combined (see Appendix II: Statistical Notes, 5). Based on the three-by-three tables of agreement, you should calculate positive and negative agreements along with 95% two-sided confidence intervals, and percents of different results by the new system for the equivocal results by the old system for each site separately and for all sites combined. From the three-by-three table, you should calculate the percent of equivocal (and/or re-test) results reported by the new system that are discordant with the equivocal results by the old system. For evaluation of the percents of different results by the new system for the equivocal results by the old system that reflect the natural variability between two measurements by the old system, you should present an allowable total difference zone and perform calculations described in Section VI.B.3.c, below (and in Appendix II: Statistical Notes, 7 and 10).

4. **Acceptance Criteria for Qualitative Assay Migration Studies**

In addition to the acceptance criteria for performance of the new system at low levels of analyte and for seroconversion panels (if applicable), we recommend you apply the following criteria to demonstrate that there are no changes to performance characteristics that could affect the safety and effectiveness of the device:

- The systematic difference between numeric values (e.g., signal output, S/CO) of the new and old systems should be either not clinically significant and not statistically significant, or, if statistically significant, should not be clinically significant (see Appendix II: Statistical Notes, 8).

- The ratio of standard deviations in the precision studies (reproducibility and within-laboratory precision) of the old and new systems should be either not

clinically significant and not statistically significant or, if statistically significant, should not be clinically significant (see Appendix II: Statistical Notes, 2 and 8).

- Conceptually, the new system measurements can be considered comparable to the old system measurements if the new system measurements of a sample are similar to the repeated evaluations of the same sample when run on the old system. For the old system measurements, one can expect high agreement between repeated measurements for the samples with high concentrations of analyte giving measurement results far from the cutoff, and some degree of discordance for the samples with concentrations measuring close to the cutoff.

  In the study using comparison panels, the lower limits of the 95% two-sided confidence intervals for the positive and negative percent agreements between the new system and the old system should be higher than 90% (see Appendix II: Statistical Notes, 9). Discordant results between the Old and new systems can only occur with samples close to the cutoff and not with moderate or high positive samples by the old system (similarly, not with moderate or low negative samples by the old system).

- You should investigate any statistically or clinically significant differences found between the two systems (e.g., in precision testing performance or any systematic difference), and perform risk assessments to determine the percentage of the intended use population that would be affected by such a difference.

B. Migration Studies for Quantitative Assays

1. **Analytical Studies for Quantitative Assays**

If you believe that some of the studies described in this section do not apply to your particular device, you should describe your reasoning in detail in your application to FDA. If the design of the analytical studies conducted for the old system were different from those described in this guidance, please contact FDA.

We recommend that you use fresh clinical specimens for all analytical studies. If this is impractical in some cases you may substitute or supplement fresh clinical specimens with archived samples. If archived samples are not available, spiked or diluted clinical samples may be used. In some instances, use of otherwise contrived matrix-specific samples may also be appropriate; however these should mimic clinical specimens as much as is feasible. We recommend that you contact FDA if you wish to discuss appropriate sample types for these evaluations. The matrix of any of these alternative specimens should be the same as that specified by the intended use of the old system.

a) Performance at Low Analyte Levels

For assays that were previously approved or licensed with a specified LoB (limit of blank) and LoD (limit of detection), the same evaluations should be repeated with the

new system. The study should demonstrate that the LoB and LoD are very similar for both systems (a protocol is described in CLSI document EP17). Specifically, the sample with a concentration at the LoD (reported as "analyte detected" approximately 95% of the time, measured by the old system) should also be reported as "analyte detected" approximately 95% of the time, if measured by the new system. See Figure 6 below regarding the LoD.

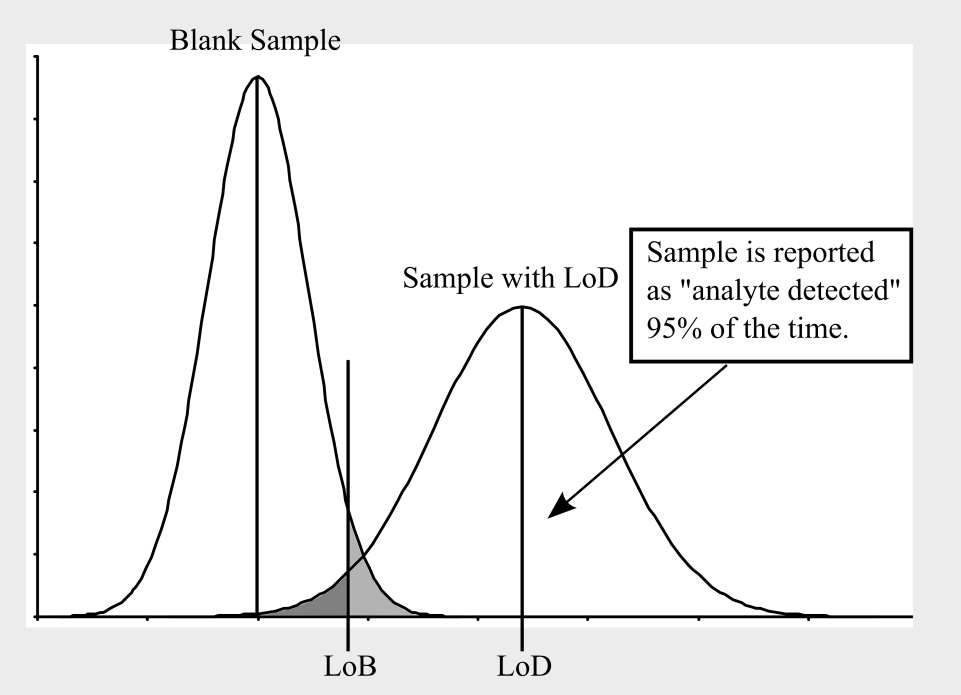

Figure 6. Relationship between the signal distributions from multiple measurements of a blank sample and a sample with an analyte concentration at the Limit of Detection.

The limit of quantitation (LoQ, or lower limit of measuring range) of the new system should be estimated and compared with the LoQ of the old system (see CLSI document EP17) and should be similar to that of the old system. The specification criteria for the LoQ of the new system should be the same as for the old system. We also recommend that the LoQ correspond to an analyte concentration level used in the precision studies.

b) Precision

This section describes precision study designs including panel composition.

(i) Composition of Precision Panel

We recommend you evaluate samples with the following levels of analyte:

- Lowest limit of the measuring range

- A level of analyte below the medical decision point
- Around the medical decision point
- A level of analyte above the medical decision point
- Upper limit of the measuring range.

In addition, you should run the appropriate control material and calibrators associated with the test kit in the precision study.

If the assay has more than one medical decision point, then samples with concentrations around these medical decision points should be evaluated. It is understood that some assays will not have a specific medical decision point, but rather a range of values; in such cases, the panel should contain samples scattered throughout the measuring range of the assay.

(ii) Within-Laboratory Precision Study

We recommend that you conduct in-house within-laboratory precision studies (to supplement the external site reproducibility studies described below in Section VI.B.1.b.iii). When appropriate and justified, the in-house within-laboratory precision study may not be called for, for example, (1) if the manufacturer established that the new system only needs to be recalibrated at relatively long time intervals (e.g., 6 months or more) and any other concerns can be appropriately addressed by the reproducibility study, or (2) if the new system is recalibrated daily, so that calibration cycle variability is inseparable from day-to-day variability (which is assessed by the reproducibility studies described below) and any other concerns can be appropriately addressed by the reproducibility study.

It may be sufficient to perform within-laboratory precision studies only on the new system. However, if the study design or composition of the precision panel of the old system precision study was very different from that described in this guidance, it may be important to perform the precision study on the old system as well. The within-laboratory precision study described below is based on CLSI document EP05.

Sources of variability for the within-laboratory precision study which we recommend you include are at least 12 days of testing, with 2 runs per day, and 2 replicates of each sample per run. These 12 days are not necessarily consecutive. If system calibration variation composes a significant portion of the total system variation, or the calibration interval fits within the time frame of the studies, then at least 2 calibration cycles should be included in the study design; otherwise you should provide your rationale for not including an evaluation of the variation introduced due to instrument calibration. For each cycle, you should include days at the beginning and end of the cycle (e.g., 3 days at the beginning and 3 days at

the end of each cycle, for each cycle). You should include other additional sources of variability in the design of the study if they are important to the specific assay (e.g., operators, lots, etc.). In such cases overall modification to the variables might be possible (e.g., spreading days of testing between different operators). If analytical and clinical performance is similar across all matrices that are included in the Intended Use of the old system, then establishing performance of the new system using the most commonly employed matrix may suffice.

(iii) Reproducibility Study

We recommend you perform the reproducibility study based on CLSI documents EP05 and EP15 on the new system. The panel composition and analyte levels for this study should be the same as described for the within-laboratory precision study above (Section VI.B.1.b.ii). We recommend that sources of variability should include testing for at least 5 days, 2 runs per day, with 3 replicates of each panel member per run at 3 laboratories (1 in-house and 2 external sites). Other sources of variability might be applicable if relevant to the specific assay (e.g., operators, etc.). If analytical and clinical performance is similar across all matrices that are indicated for the old system, then establishing performance of the new system using the most commonly employed matrix may suffice.

For each concentration level, similar information should be available for the old system. If this is not the case, a new reproducibility study for the old system should be performed with study design and concentration levels as described in this section.

c) Linearity Study

We recommend that you evaluate linearity for the new system according to CLSI document EP06. The degree of linearity can be quantified using the maximum deviation from linearity (i.e., the delta described in CLSI document EP06). Your linearity study results should demonstrate that the delta of the new system is not greater than the observed delta in the linearity studies of the old system. You should determine the appropriate number of replicates in the linearity study for the new system based on the precision studies of the new system.

2. Comparison Studies

You should perform comparison studies using comparison panels. Relatively minor changes to the old system might not warrant all comparison studies. The extent of the utility of these studies can be evaluated on a case-by-case basis in consultation with the FDA.

a) Comparison Panels

For each analyte, the composition of the quantitative assay comparison panels should consist of at least 180 samples[9], including the following:

- 150 of the 180 samples should span the measuring range of the assay with approximately equal numbers of samples at low, medium and high analyte concentrations.

- If the performance of the assay at low level concentrations is clinically important, the remaining panel members should consist of at least 30 samples (or about 15%-20% of the total) that should be patient samples with concentration at or near zero. The zero-level samples may need to consist of different matrices and/or be from persons with unrelated medical conditions.

- It is preferable to use non-contrived clinical samples. However, where clinical samples are not available, or volumes are low, pooling is a possible strategy. If pooling would not be efficacious, then dilutions, made serially of individual samples with a high analyte concentration diluted into appropriate clinical matrices, can be used. Spiked samples are acceptable, if they are prepared from individual samples with high analyte concentrations in clinical matrices (i.e., the same sample should not be used repetitively).

- Archived samples are acceptable if stored in accordance with the package insert instructions. Samples should be masked, and randomized with respect to the order in which they are run.

- The panel should consist of appropriate analyte-specific members. For example, different subtypes or strains of an infectious agent should be included when applicable.

b) Comparison Study Design

You should test the comparison panels on the old system at a minimum of one site. This may be done in-house. However, you may want to use more than one old system to better assess instrument bias. The new system should be tested at a minimum of three sites (one may be in-house) with at least one reagent kit lot. Each panel member should be tested at least four times: once with the old system and three times with the new system. You should send the same positive and negative panels to each site, rather than dividing the panel between the three sites. Three different builds of the new system should be tested, one at each of the three sites.

3. Statistical Analysis of Data

[9] For more information on sample size, see Appendix II: Statistical Notes, 7.

a) Within-Laboratory Precision Study

- For each sample tested, we recommend you present the mean value with variance components (standard deviation and percent CV) for the new system.

- For each analyte concentration, you should present similar information for the old system. This information can be taken from the precision study originally used to demonstrate the performance of the old system if the study design and the analyte concentration levels are comparable to those used to assess the new system. The information about variance components (standard deviation and percent CV) can be obtained from the precision study found in the old system labeling. If the study design and analyte concentrations are different than those described in this guidance document, you should perform a new precision study for the old system.

- For each analyte concentration, (using precision profiles of the New and old systems) we recommend you provide 95% confidence intervals for repeatability and within-laboratory precision (standard deviation and percent CV) of the New and old systems In addition, we recommend you provide the ratio of the repeatability standard deviations of the New and old systems and the ratio of within-laboratory standard deviations along with the 95% confidence interval for these ratios. The confidence interval can be based on the F-statistic for a ratio of variances.

b) Reproducibility Study

- For each sample tested, you should present the mean value with variance components (standard deviation and percent CV) for the new system for each site separately and for all sites combined.

- For each analyte concentration level, you should present similar information for the old system. This information can be taken from the precision study originally used to demonstrate the performance of the old system if the study design and the analyte concentration levels are comparable to those used to assess the new system. The information about variance components (standard deviation and percent CV) can be obtained from the precision study found in the old system labeling. If the study design and analyte concentrations are different than those described in this document, a new precision study for the old system should be performed.

- For similar concentration levels, you should provide repeatability (within-run precision), within-laboratory precision (standard deviation and percent CV) and reproducibility separately and combined over three sites for the new systems and old system (based on precision profiles of the new system and old system). In addition, you should provide the ratio of repeatability standard deviations along with 95% confidence intervals and the ratio of the reproducibility standard deviations of the New and old systems. The

confidence intervals for the ratios can be based on the F-statistic for a ratio of variances (see Appendix II: Statistical Notes, 2).

c) Comparison Panels

- You should provide a scatter-plot of numeric values of the old system (X-axis) vs. numeric values of the new system (Y-axis) with descriptive statistics for data at each site, as well as for all sites combined. Both axes should have the same scale and the line of identity (y=x), and the same scale on the axes should be applied to the data from each site.

- You should provide an estimation of systematic differences between absolute numeric values by the New and old systems (CLSI document EP09) for each site and for all sites combined. You should perform an appropriate regression analysis (such as a Deming regression) which accounts for the random errors associated with the Old and new system measurements, and provide the 95% confidence intervals of the slope and intercept from this analysis. We recommend that, for the data from each site and for the combined site data, you draw the regression line on the corresponding scatter plots and plot the fitted lines (for each site and all sites combined) on the same corresponding figures. Using the regression equation, you should calculate the systematic bias at all medically important points along with 95% confidence intervals.

- Conceptually, the new system measurements can be considered comparable to the old system measurements if the new system measurements of a sample are similar to the repeated evaluations of the same sample when run on the old system. Using the reproducibility results of the old system, one can construct limits or boundaries that define where 95% of the differences between two repeated measurements by the old system are inside of these limits (see Appendix II: Statistical Notes, 10). These limits define an Allowable Total Difference (ATD) zone (see CLSI document EP21) for differences between new system and old system measurements. It is anticipated that no less than 95% of the sample results will fall within the ATD zones (see Figure 7, below). For details see Appendix II, Statistical Notes, 11.

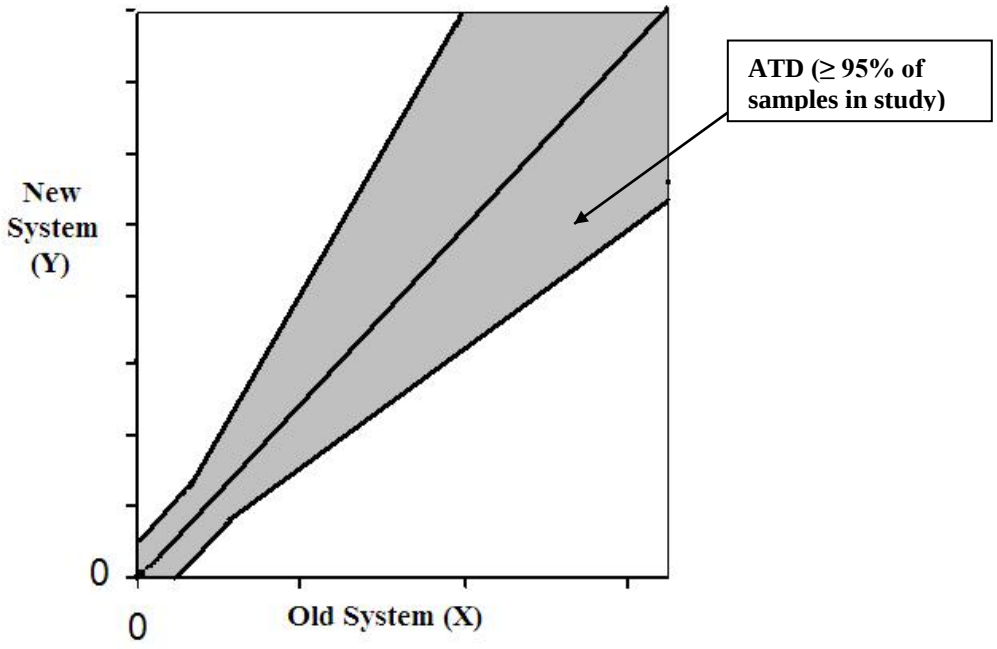

Figure 7. Allowable Total Difference Zone.

- You should calculate (i) the percentage of the samples at low, medium, and high concentration ranges that fall within the ATD zone at each site and averaged over three sites, (ii) the percentage of samples over the entire range that fall within the ATD zone with a lower limit of 95% one-sided confidence interval at each site, and (iii) the percentage of samples over the entire range that fall within the ATD zone averaged over three sites using a bootstrap technique, you should provide the 95% one-sided confidence interval for the percentage of the samples over the entire range that fall within the ATD zone averaged over three sites.

4. Acceptance Criteria for Quantitative Assay Migration Studies

In addition to the acceptance criteria for LoB, LoD, LoQ (see section VI.B.1.a) and linearity (see section VI.B.1.c), we recommend that you apply the following criteria:

- The systematic difference between numeric values of the new and old systems should be either not clinically significant and not statistically significant, or if statistically significant, it should not be clinically significant. (See Appendix II: Statistical Notes, 8).

- The ratio of standard deviations in the precision studies (reproducibility and within-laboratory precision) of the old and new systems should be either clinically and statistically insignificant, or if statistically significant it should not be clinically significant. (See Appendix II: Statistical Notes, 8).

- The percentage of the observations over the entire range that fall within the ATD zone should be close to 95% with a lower limit of the 95% one-sided confidence interval higher than 90%. The percentages of the observations that fall within the ATD zone for the low, medium, and high ranges should approach 95% for each range. (See Appendix II: Statistical Notes, 11).

- If applicable, the percent of positive results by the new system among zero-level samples should be consistent with a Type I error of LoB of the old system across three sites. (Type I error is the probability of having True Negative samples, i.e., those with zero analyte concentration, give values that indicate the presence of analyte. Usually, Type I error is set as 5% or less).

- You should investigate any differences found between the two systems and perform a risk assessment to determine what percentage of the intended use population would be affected by these differences.

C. Migration Studies for Semi-Quantitative Assays

Semi-quantitative assays are those which are neither qualitative nor quantitative. An example of a semi-quantitative assay is one with a few ordinal categories (negative, trace, +, ++, +++) where the order of categories contains information used during the interpretation of the assay results. Another example of a semi-quantitative assay is an assay which has continuous numeric values but does not have a linearity property, and, therefore is not quantitative. The types of the analytical and comparison migration studies depend on the nature of the semi-quantitative assay. If your semi-quantitative assay has multiple cutoffs, the performance of the new and old system should be evaluated near each cutoff separately (e.g., precision and systematic differences at each cutoff). You should contact the appropriate FDA CDRH or CBER Division to obtain recommendations about migration studies for your particular semi-quantitative assay.

D. Migration Studies for Point Of Care (POC) Assays

Before preparing to use the migration studies approach for a POC assay, you should determine whether the assay is quantitative, qualitative or semi-quantitative, according to the definitions in this guidance. The migration studies should be conducted in the hands of intended users of the POC assays. You should contact the appropriate FDA CDRH or CBER Division to obtain recommendations about migration studies for a POC device.

VII. Other Studies

Depending upon the unique characteristics of the qualitative, quantitative or semi-quantitative assay being migrated to the new system, the following studies may be called for. *If not previously conducted for the old system, they should be performed for the new system*.

If you determine that a study described below is not applicable for your system, you should describe your reasoning in detail in your application to FDA. FDA will consider such explanations on a case-by-case basis particularly for manual to semi-automated or automated system migrations.

- Carry-over or cross-contamination studies: The importance of repeating these studies on a new system can be ascertained by a thorough analysis of the new system. As mentioned in Section V of this document, block diagrams and side-by-side comparison tables would be beneficial in this determination. Changes to specific physical features such as a change in sample pipettor design or the layout of the new system could indicate the need for new carry-over studies. If a carry-over study for the new system is appropriate and the new design is sufficiently similar to the old system, the new study can be the same as previously used for the old system. Samples with high positive concentrations of analyte should be tested alternating with analyte-negative samples in patterns dependent on the operational function of the instrument. The concentration of analyte in the high positive samples should exceed 95%-99% or more of the results normally obtained in clinical samples from diseased patients in the intended use population. This testing should be done over multiple runs (at least 5 runs are recommended by the Commission of European Communities[10]).

- Matrix equivalency and recovery studies: Presumptively, because there are no changes to the assay, there should be no new effects on the assay performance due to different matrices unless physical alterations of the new system could create such an effect. Similarly, recovery studies should not be affected by the migration of an assay from an old to a new system.

- Interfering substances studies: Presumptively, with no changes to the assay, there should be no new effects on the assay performance due to interfering substances.

- On-board reagent/calibrator and sample stability studies: Unless there are physical or process changes to the new system, presumptively, there should be no effect on assay performance due to on-board reagent or sample stability.

- Cross-reactivity studies: Presumptively, as there are no changes in assay components, a migration to a new system should have no effect on the existing cross-reactivity.

- Hook effect studies: Presumptively, because there is no assay change, the parameters of the hook effect should be the same, unless physical alterations in the new system could create such an effect.

---

10 Commission Decision 2002/364/EC of 7 May 2002 on common technical specifications for in vitro-diagnostic medical devices [Official Journal L 131 of 16.05.2002].

- Verification of kit control material and calibrators: Presumptively because there is no change to the assay, the control results and calibration ranges should remain the same, unless physical alterations in the new system could create new effects.

- If new analyte-specific information requires new studies to bring the assay's performance characteristics up-to-date, please contact FDA for guidance.

VIII. Molecular Assays

There are specific criteria that are unique to nucleic acid tests (NAT) and therefore NATs present additional specific concerns over serological and antigen assays:

- When appropriate, you should provide testing with panels showing a rise in viral titer over time from serial bleeds (viremic profile). Similar to seroconversion panels, they should have a minimum number of days between bleeds and begin with at least one negative bleed. They should be of clinical relevance to the appropriate individual marker.

- Carryover studies: Because of the increased risks of carryover due to the amplification methodologies utilized in molecular testing (e.g., PCR, TMA), you should perform carryover studies for all NAT migration studies.

- Sample stability: Because of the nature of DNA and especially RNA, careful attention should be given to the stability of samples in relation to on-board storage and manipulation.

- Sample processing: The processes of purification and extraction of DNA or RNA from clinical samples is critical to the success of molecular tests. You should evaluate any additions or modifications associated with the new system that could affect these processes.

- Validation of control material and calibrators: You should perform these studies on the new system due to the sensitive nature of molecular assays.

- For molecular assays detecting multiple analytes, please contact FDA for further discussion.

IX. Regulatory Outcomes

- Should the acceptance criteria noted in Sections VI.A.4 or VI.B.4 above be met, it would be appropriate for the sponsor to claim that the new system does not compromise the results as compared to the old system. It would not be appropriate to claim improved performance characteristics. It would also not be appropriate to compute clinical performance claims for the new system based on the migration

studies described here, since these studies are analytical, rather than clinical. Should you wish to develop more extensive claims, the migration studies paradigm would not be an appropriate scientific approach.

- If the acceptance criteria are not met and the aberrant performance could affect clinical management, you will be asked to perform a complete clinical study presenting the clinical performance of the assay on the new system.

X. Glossary

For the purposes of this document the following definitions are used. HTD referenced terms are based on the CLSI Harmonized Terminology Database.

C_5: a concentration at which repeated tests of a sample with this concentration under stipulated conditions are 95% negative (or 5% positive) (see CLSI EP12-A2).

C_{95}: a concentration at which repeated tests of a sample with this concentration under stipulated conditions are 95% positive (see CLSI EP12-A2).

C_{50}: a concentration at which repeated tests of a sample with this concentration under stipulated conditions are 50 % positive (or 50% negative). Under ideal circumstances, C_{50} will exactly equal the cutoff established by the manufacturer.

Calibrators: a substance or device that is based on a reference preparation or in which the analyte concentration or other quantity has been determined by an analytical procedure of stated reliability. Calibrators are used to calibrate, graduate, or adjust a measurement [HTD].

Carry-over: amount of analyte carried by the measuring system from one sample reaction into subsequent sample reactions, thereby erroneously affecting the apparent amounts in subsequent samples [HTD].

Control material: a device, solution, or preparation intended for use in the quality control process to monitor the reliability of a test system and to evaluate its performance within established limits.

Cross-reactivity: the ability of a drug, metabolite, a structurally similar compound other than the primary analyte, or even unrelated compound to affect the assay [HTD].

Cutoff value (CO): for a qualitative test, the threshold above which the result of the test is reported as positive and below which the result is reported as negative.

High negative sample (C_5): a sample with a concentration of analyte close to the C_5 as determined by the old system. This term is equivalent to a "weak negative sample" for example as used in the guidance document "Recommendations for CLIA Waiver Applications for Manufacturers of IVDs", http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070890.pdf.

Hook effect (high dose hook effect): effect of a suboptimal antigen-antibody reaction in which either the antibody or antigen is in excess resulting in an incomplete, or blocked reaction leading to a decreasing signal response at very high levels of analyte. It is used interchangeably with "prozone effect," [HTD].

Interfering substances: endogenous (e.g., blood components, acidic polysaccharides) or exogenous (e.g., talc, anticoagulant) substances in clinical samples that can cause false-positive or false-negative results in a test system [HTD].

Limit of blank (LoB): highest measurement result that is likely to be observed (with a stated probability) for a blank sample (a sample with concentration at or near zero) (CLSI EP17-A; [HTD]).

Limit of detection (LoD): the lowest concentration of analyte that can be reported to be present at a specified level of confidence, although perhaps not quantified to an exact value. Similarly, an amount of analyte in a sample for which the probability of falsely claiming the absence is β (type II error) given a probability α (type I error) of falsely claiming its presence (CLSI EP17-A; [HTD])

Limit of quantitation (LoQ): the lowest amount of analyte in a sample that can be quantitatively determined with {stated} acceptable precision and {stated, acceptable} accuracy, under stated experimental conditions (CLSI EP17-A; [HTD]).

Linearity: ability (within a given range) to provide results that are directly proportional to the concentration (amount) of the analyte in the test sample [HTD].

Linearity studies: studies to determine the analyte concentration range over which the testing systems results are acceptably linear.

Low positive sample (C_{95}): a sample with a concentration of analyte close to C_{95} as determined by the old system. This term is equivalent to "weak positive sample" for example as used in the guidance document "Recommendations for CLIA Waiver Applications for Manufacturers of IVDs", http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070890.pdf.

Measurand: particular quantity subject to measurement. The term "measurand" and its definition encompass all quantities while the commonly used term "analyte" refers to a tangible entity subject to measurement [HTD].

Measuring interval (measuring range): set of values of measurands for which the error of a measuring instrument is intended to lie within specified limits. The range of values (in units appropriate for the analyte [measurand]) over which the acceptability criteria for the method have been met; that is, where errors due to nonlinearity, imprecision, or other sources are within defined limits (CLSI EP06-A, [HTD]).

Medical decision level (medical decision point): a level or concentration at which a test is interpreted for patient care and treatment.

Moderate positive sample: a sample with a concentration close enough to the cutoff and at which one can anticipate positive results by the old system approximately 100% of the time.

Negative percent agreement: the proportion of samples negative by the old system for which the results by the new system are negative (see FDA guidance titled "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests" at

http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf).[11]

New system: an unapproved/unlicensed system (assay, instrument, and software) to which the assay is migrating from a previously approved/licensed system.

Old system: an approved/licensed system (assay, instrument and software) from which the assay is migrating to a currently unapproved/unlicensed system.

Positive percent agreement: the proportion of samples positive by the old system for which the results by the new system are positive. (See FDA guidance "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests", http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf).

Qualitative assay: assay in which nominal properties of measurand are identified.

Quantitative assay: assay in which the amount or concentration of an analyte is measured and expressed as a numeric quantity value in appropriate measurement units.

Repeatability: closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement (See CLSI EP05 [HTD]).

Reproducibility: closeness of agreement between the results of measurements of the same measurand and carried out under changed conditions of measurement. Reproducibility conditions are conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment and can include additional variables such as days, replicates, and runs (See CLSI EP05 [HTD]).

Risk analysis: systematic use of available information to identify hazards and to estimate the risk. Risk analysis includes examination of different sequences of events that can produce hazardous situations and harm [HTD].

Semi-quantitative assay: assay which is neither qualitative nor quantitative (such as an assay with a few ordinal categories; negative, trace, +, ++, +++).

Spiked sample: a clinical sample to which exogenous analyte has been added to create specified levels of signal.

Systematic difference: a mean of the measurand on the new system minus a value of the same measurand as performed on the old system that would result from an infinite number of measurements carried out under the stipulated condition (based on HTD).

Within-laboratory precision: precision over a defined time and operators, within the same facility and using the same equipment; calibration and reagents may vary. Formerly, the term "total precision" was used in CLSI EP05 [HTD].

[11] The general definition from the cited guidance is adapted in this guidance since the cited guidance refers to clinical subjects, whereas this guidance does not involve subjects, but rather specimens or samples. In addition, the term "non-reference standard" in the cited guidance is analogous to "old system" in this guidance; the term "test" in the cited guidance is analogous to "new system".

XI. FDA Guidance Documents and CLSI Guidelines Referenced

1. *Quality System Information for Certain Premarket Application Reviews; Guidance for Industry and FDA Staff* issued 2003, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070897.htm.
2. *Guidance for Industry and FDA Staff: Replacement Reagent and Instrument Family Policy* issued 2003, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079185.htm
3. *Guidance for Industry and Staff: Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices* issued 2005, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm.
4. *Guidance for Industry and FDA Staff: Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices* issued 2008, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079632.htm.
5. *Guidance for Industry and FDA Staff: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests* issued 2007, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.

Clinical and Laboratory Standards Institute (CLSI) documents:

1. CLSI. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition*. CLSI document EP05-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2004.
2. CLSI. *Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline*. CLSI document EP06-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2003.
3. CLSI. *Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline – Second Edition (Interim Revision)*. CLSI document EP09-A2-IR. Wayne, PA: Clinical and Laboratory Standards Institute, 2010.
4. CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline-Second Edition*. CLSI document EP12-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
5. CLSI. *User Verification of Performance for Precision and Trueness; Approved Guideline – Second Edition*. CLSI document EP15-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2006.

6. CLSI. *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline – Second Edition*. CLSI document EP17-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.
7. CLSI. *Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline*. CLSI document EP21-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2003.

Appendix I - Migration Studies for Blood Donor Screening Assays[12]

1. Introduction

Blood Donor Screening Assays for infectious agents, reviewed under Biologics License Applications (BLA), are generally held to stringent standards of sensitivity and specificity. Typically, clinical studies for licensure of products such as HIV assays involve testing of over 1000 known positives and 6000 to 10,000 low risk individual samples (or pools) collected from the intended use population. Consequently, FDA recommends larger study sizes for migrating blood screening assays to new systems.

Otherwise, except as specifically noted below, the same considerations apply to blood screening assays as described for qualitative assays in Section VI.A of this document, "Migration Studies for Qualitative Assays."

a. Performance at Low Analyte Levels

For immunoassays, FDA recommends that at least 20 seroconversion panels, or as many as are available (whichever is less) be studied, comparing the new and the old systems head to head. For nucleic acid tests (NAT), FDA recommends the head-to-head testing of as many seroconversion panels as were tested for licensure of the old system (typically 10). Both qualitative results and numeric signal (e.g. signal output, S/CO ratios) should be compared.

b. Precision Study (Within-Laboratory Precision)

FDA recommends that sponsors compare the new and old systems in a Precision Study as outlined in Section VI.A.1.b.

c. Reproducibility Study

FDA recommends that sponsors compare the new and old systems in a Reproducibility Study essentially as outlined in Section VI.A.1.b.iii. However, FDA recommends including in the panel at least one truly negative sample (i.e. not just a positive sample diluted to below the cutoff), and that testing be performed at three sites, one of which may be in-house.

2. Comparison Panels

a. Positive Panel Members

A positive comparison panel should consist of approximately 100 positive samples, to include 60-80 specimens with signals $\leq$3X the cutoff for an immunoassay or analyte concentrations $\leq$3X the 95% LoD for a NAT. Authentic clinical specimens are preferred: however, specimens may be diluted to the recommended range in negative matrices (individual, not pooled) if clinical specimens are not available. The panel should be tested on the new system at three sites, one of which can be in-house. This panel should

[12] Appendix 1 does not apply to immunohematology tests licensed by CBER.

also be tested at least once on the old system (this can be in-house). We recommend use of at least two of the old systems to better assess systematic differences between old systems. The data should be analyzed by S/CO regression and analysis of bias using scatter-plots or similar graphical presentations, as described in Section VI.A.3.c.

b. Negative Panel Members

Option 1: A negative comparison panel should consist of 3000 known negative samples (or pools), or specimens obtained from a low risk study, with adequate follow-up. The panel should be tested on the new system, distributed over three sites, one of which may be in-house (e.g., 1/3 at site one, 1/3 at site two & 1/3 at site three). The data should be analyzed for agreement of the point estimate of specificity (with the 95% confidence interval) for the new system with the original point estimate of specificity (with the 95% confidence interval) from the original trial of the old system.

Option 2: As an alternative option, the sponsor may wish to test head to head on the old and new systems at three external sites the highest 10% of negative specimens from the original clinical trial of the assay on the old system, if the specimens have been stored under conditions defined in the instructions for use of the assay.

3. Acceptance Criteria

Except as noted above, the same acceptance criteria as recommended in Section VI.A.4 also apply to blood donor screening assays.

Additionally, sponsors may recommend statistical analysis protocols based on estimating false positive and false negative ratios from the negative and positive comparison panel studies, respectively. Sponsors interested in this approach should determine an appropriate model for the S/CO distribution of each panel (positive or negative) together with a proposed method of analysis.

4. Interfering Substances and Conditions

Only substances and conditions that represent a reasonable risk of interference in the new system should be studied. For instance, interfering conditions such as hemolysis or hyperlipidemia might influence pipetting or washing steps and should be included in migration studies. Conversely, it would seem unlikely that cross-reactivity of, for instance, an HIV NAT assay with HTLV would likely be influenced by migration to a new system.

Each interfering substance/condition may be tested in-house using a panel of approximately 10 Low Positives with signals $\leq$3X the cutoff for an immunoassay or analyte concentrations $\leq$3X the 95% LoD for a NAT. (These studies can be performed by spiking the analyte at the level described above in negative specimens (for the analyte) with interfering substances/conditions).

Each interfering substance/condition may be tested in-house using a panel of approximately 10 True Negatives.

Because migration studies extend the life of individual assays and new information becomes available over time, it is important to update the performance characteristics of the assay during migration of the assay. Based on the type of data considered for the original assay,

examples of new information that may trigger additional study are newly recognized potentially interfering substances or conditions, or new variants, or new recombinants.

Appendix II - Statistical Notes

1. Evaluation of C_5 and C_{95} using Precision Studies

Consider that an assay (qualitative or quantitative) has a numeric output in which higher values indicate higher levels of the analyte and analytical variation about the C_{50} is approximately normal (or approximately normal after appropriate transformation). If the standard deviations (SD) in the precision studies of the old system for concentrations around the cutoff value are almost constant, then:

$C_{95} = C_{50} + 1.645 \times SD$, and
$C_5 = C_{50} - 1.645 \times SD$

For example, if the cutoff for optical density (OD) value is 1.00 and the SD around the cutoff is approximately 0.10, then C_{95} is approximately 1.16 OD (=1.00+1.645 x 0.10) and C_5 is approximately 0.84 OD (=1.00 -1.645 x 0.10). In other words, a sample with an actual OD of 1.16 produces positive results (above 1.00) approximately 95% of the time and a sample with an actual OD of 0.84 produces negative results (below 1.00) approximately 95% of the time.
If the coefficient of variation (CV) in the precision studies of the old system for concentrations around the cutoff value are almost constant, then
$C_{95} = C_{50} + 1.645 \times CV \times C_{95}$ and $C_5 = C_{50} - 1.645 \times CV \times C_5$. From here,

$C_{95} = C_{50} / (1 - 1.645 \times CV)$ and
$C_5 = C_{50} / (1 + 1.645 \times CV)$.

For example, if the cutoff has an OD value of 1.00 and the %CV around the cutoff is approximately 10% (i.e., CV=0.10), then C_{95} is approximately 1.20 OD and C_5 is approximately 0.86 OD.

If the limit of blank (LoB) is used as a cutoff, then the concentration C_{95} is the same as the limit of detection (LoD) and zero concentration is C_5 (see CLSI EP17).

2. Calculating Confidence Interval for the Ratio of the Standard Deviations

Let σ^2_{New} and σ^2_{Old} be true variances and SD^2_{New} and SD^2_{Old} be estimated variances. Consider the expression $F = [SD^2_{New}/ \sigma^2_{New}] / [SD^2_{Old}/ \sigma^2_{Old}]$ which follows an F distribution with degrees of freedom df=ν_1 (numerator) and df=ν_2 (denominator): $F(\nu1, \nu2)$.[13] For a level of confidence α, let $F(1-\alpha/2, \nu1, \nu2)$ and $F(\alpha/2, \nu1, \nu2)$ be critical values that

Prob $\{F(\alpha/2, \nu_1, \nu_2) < [SD^2_{New}/ \sigma^2_{New}] / [SD^2_{Old}/ \sigma^2_{Old}] < F(1-\alpha/2, \nu_1, \nu_2)\} = 1-\alpha$;

[13] For more details see Searle S.R., Cassella G. and McCullagh C.E. Variance Components (1992) Wiley&Sons; Hahn G.J., Meeker W.Q. Statistical Intervals. A Guide for Practitioners (1991) Wiley&Sons.

⇨ Prob $\{(1/F(1-\alpha/2, \nu_1, \nu_2)) * (SD^2_{New} / SD^2_{Old}) < \sigma^2_{New} / \sigma^2_{Old} < (1/F(\alpha/2, \nu_1, \nu_2)) * (SD^2_{New} / SD^2_{Old})\} = 1-\alpha$;

Therefore, the interval from $(1/F(1-\alpha/2, \nu_1, \nu_2)) * (SD^2_{New} / SD^2_{Old})$ to $(1/F(\alpha/2, \nu1, \nu_2)) * (SD^2_{New} / SD^2_{Old}\}$ is a two-sided 95% confidence interval for $\sigma^2_{New} / \sigma^2_{Old}$. To find a confidence interval for the ratio of standard deviations $\sigma_{New} / \sigma_{Old}$, take square roots. Note that the F distribution has the property that $F(1-\alpha/2, \nu_1, \nu_2) = 1/F(\alpha/2, \nu_2, \nu_1)$.

For determination of degrees of freedom, you should refer to CLSI EP05. For example, for a precision study conducted with 3 sites, 5 days per site, 2 runs per day and 3 replicates per run, the estimate of SD of repeatability for one run has 2 degrees of freedom (=3-1) and the estimate of SD of repeatability for all thirty (=3*5*2) runs has 60 degrees of freedom (=3*5*2*2).

3. Deming and Passing-Bablok Regressions

Ordinary least squares regression analysis supposes that one of the methods is without random measurement error, and the standard deviation of the error distribution of the other method is constant through the measurement range. An assumption about the absence of a random measurement error is not fulfilled in the migration comparison study data because both methods, X and Y, are subject to random measurement errors; therefore, Deming and/or Passing-Bablok regression analyses are recommended. An assumption of a constant standard deviation of X and Y methods through the entire measurement range does not hold for most of the assays. For example, very frequently, the standard deviation increases as the concentrations increases. In situations like this, weighted Deming regression can often be used.

The analysis of the numeric values in the migration comparison study data should include an appropriate type of Deming regression (e.g., weighted) for each site separately and for all sites combined.

To apply a Passing-Bablok regression analysis, the comparison study data should adhere to the following requirements: $CV_Y / CV_X = \beta$ where β = the slope (it is assumed that X and Y have approximately the same random measurement errors). In an article by Passing, H. and Bablok, W.[14], the authors performed simulations to investigate how inequality in the random measurement error of X and Y affects Passing-Bablok regression. The authors found that if $0.67<CV_Y/CV_X<1.5$ then Passing-Bablok regression is acceptable. If $1.5<CV_Y/CV_X<2.5$ (or $0.4<CV_Y/CV_X<0.67$) and precision of X is less than 7% then Passing-Bablok regression is acceptable.

When applying regression analysis of study data for all sites combined, note that only one measurement per sample is reported from the old system (X data), while an average of three measurements (one per study site) is reported for the new system (Y data). In the case of combined data, the CV of the average of Y is less because the average of Y is more precise

[14] Passing, H. and Bablok, W.[14] "Comparison of Several Regression Procedures for Method Comparison Studies and Determination of Sample Sizes" in J. Clin. Chem. Clin. Biochem., Vol. 22, 1984, pp.431-445,

than one measurement of Y. So, when combining data from all sites, the CV of the average of the Y measurements is calculated as the standard deviation of the Y data divided by the square root of 3 (CV of average Y = CV_Y * 0.58).

In addition to the Deming regression, you could also include Passing-Bablok regression analysis for each site separately. If the precision of X is less than 7% then, in addition to the Deming regression analysis for the combined data, present Passing-Bablok regression analysis for the combined data.

4. Examples of Data Tables for Qualitative Assays

The following are examples of detailed tables of agreements referred to above within Section VI.A.3.c-Comparison Panels (for Qualitative Assays).

Positive panel samples:

	Old System Positive		
	Low Positive (close to C_{95})	Moderate Positive	High Positive
New System Positive	27	30	40
New System Negative	3		
Total	30	30	40

Negative panel samples:

	Old System Negative	
	Low and Moderate Negative	High Negative (close to C_5)
New System Positive		3
New System Negative	70	27
Total	70	30

If the CO of the assay is the LoB, the columns of Low Negative and Moderate Negative can be combined as in the example above.

Table of agreement 2-by-2:

		Old System	
		Negative	Positive
New System	Negative	A_1	B_1
	Positive	A_2	B_2
Total		N_A	N_B

- Calculate Negative Percent Agreement (NPA) as A_1/N_A with 95% two-sided CI;
- Calculate Positive Percent Agreement (PPA) as B_2/N_B with 95% two-sided CI;

5. Examples of Data Tables for Qualitative Assays with Equivocal (and/or Re-testing Zone) Results.

Examples of tables for an assay with an equivocal zone of 0.8-1.2 S/CO is provided below (E_1=0.8 S/CO and E_2=1.2 S/CO).

Detailed tables of agreement:

		Old System				
		Negative		**Equivocal**	**Positive**	
		Strong and moderate negative	**Close to C_5 of E_1**	**0.80-1.20**	**Close to C_{95} of E_2**	**Strong and moderate positive**
New System	**Negative**	X	X	X		
	Equivocal		X	X	X	
	Positive			X	X	X
Total		70	60			70

		Old System		
		Negative	**Equivocal**	**Positive**
New System	**Negative**	A_1	B_1	C_1
	Equivocal	A_2	B_2	C_2
	Positive	A_3	B_3	C_3
Total		N_A	N_B	N_C

- Calculate Negative Percent Agreement (NPA) as A_1/N_A with 95% two-sided CI;
- Calculate Positive Percent Agreement (PPA) as C_3/N_C with 95% two-sided CI;
- Calculate percents of negative, equivocal and positive results by the new system among N_B equivocal results by the old system.

In addition, an example of more detailed information about PPA and NPA is shown in the format below:

		Percent	95% CI
Negative Percent Agreement	All samples negative by the Old system		
	Negative samples below C_5		
Positive Percent Agreement	All samples positive by the Old System		
	Positive samples above C_{95}		
Equivocal	Provide numbers and percents of equivocal, negative and positive results by the New System.		

6. Calculating Score Confidence Intervals for Percentages and Proportions

The following are additional recommendations for performing statistical analyses of percentages or proportions. Confidence limits for positive percent agreement and negative percent agreement can be calculated using formulas for calculating a confidence interval for a binomial proportion. There are several different methods available. We suggest that either a score method described by Altman, et al. (Altman D.A., Machin D., Bryant T.N., Gardner M.J. eds. *Statistics with Confidence.* 2nd ed. British Medical Journal; 2000) or a Clopper-Pearson Method (Clopper CJ, Pearson E. *Biometrika* 1934; 26:404-413) be used.

An advantage with the score method is that it has better statistical properties and it can be calculated directly. So with n=100 samples and 96/100=96% agreement, the score lower confidence bound is 90.2% whereas the Clopper-Pearson lower confidence bound is 90.1%. In this document, we have illustrated the reporting of confidence intervals using the score approach. For convenience, we provide the formulas for the score confidence interval for a percentage. Note that the lower bound of a two-sided 95% score confidence interval is the same as the lower bound of a one-sided 97.5% score confidence interval; and the lower bound of one-sided 95% score confidence interval is the same as the lower bound of a two-sided 90% score confidence interval.

A two-sided 95% score confidence interval for the proportion of A/B is calculated as:

$\left[100\%\left(Q_1 - Q_2\right)/Q_3, 100\%\left(Q_1 + Q_2\right)/Q_3\right]$, where the quantities Q_1, Q_2, and Q_3 are computed from the data using the formulas below. For the proportion of A/B:

$Q_1 = 2 \cdot A + 1.96^2 = 2 \cdot A + 3.84$

$$Q_2 = 1.96\sqrt{1.96^2 + 4 \cdot A \cdot (B - A)/B} = 1.96\sqrt{3.84 + 4 \cdot A \cdot (B - A)/B}$$

$$Q_3 = 2 \cdot (B + 1.96^2) = 2 \cdot B + 7.68$$

In the formulas above, 1.96 is the quantile from the standard normal distribution that corresponds to 95% confidence. For calculation of 95% one-sided score confidence interval, use 1.645 in place of 1.96 in the formulas above.

7. Allowable Total Difference For Qualitative Assays with Equivocal (and/or Re-testing Zone) Results

Based on the reproducibility of the old system for the samples close to the zone of equivocal results (and/or the re-test zone), you can establish an allowable total difference zone as presented in Figure 8 below where E_1 and E_2 are the cutoffs for the zone of equivocal results. In Figure 8, r is a result of one measurement for a sample with a result below the cut-off E_1 and repeated measurements are below E_1 with high probability; likewise R is a result of one measurement for a sample with a result above the cut-off E_2 and repeated measurements are above E_2 with high probability.

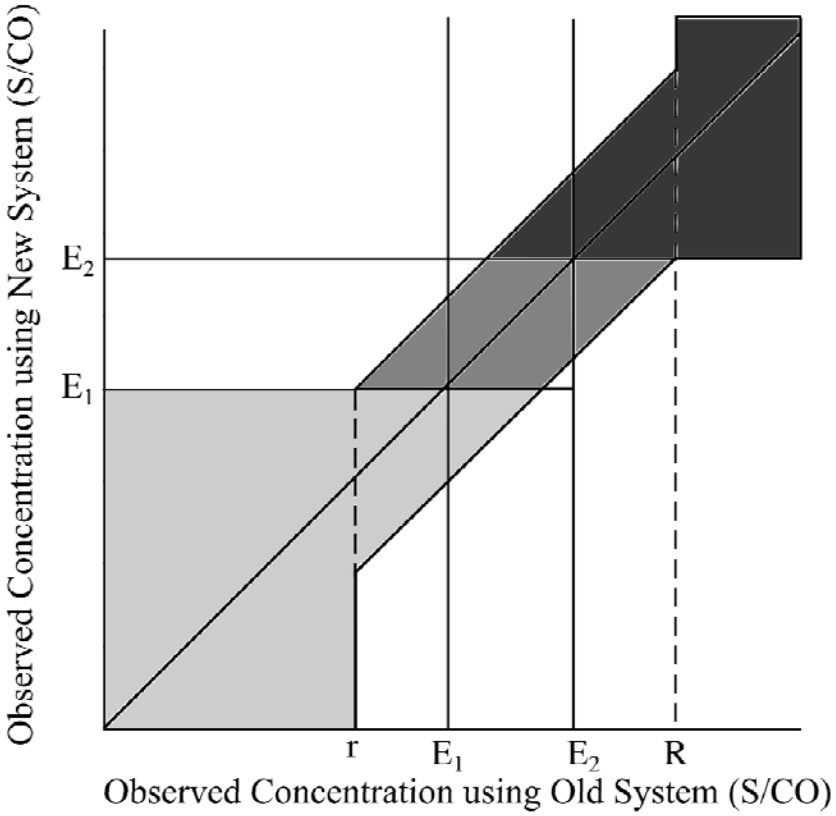

Figure 8. Allowable total difference zones for a qualitative assay with equivocal results. Light gray: Negative results by New System, Medium gray: Equivocal results by New System, Dark gray: Positive results by New System

You should calculate the percentages of the samples with Negative, Equivocal, and Positive results by the new system that fall within the five concentration ranges defined by the old system: below r, r-E_1, E_1-E_2, E_2-R, above R.

8. Relationship between Clinical and Statistical Significance

Observed Difference From *Clinical* Point of View	Observed Difference From *Statistical* Point of View	Interpretation
Not-significant	Not-significant	Acceptable[1]
Not-significant	Significant	Acceptable[1]
Significant	Not-Significant	Additional data are likely needed[2]
Significant	Significant	Not acceptable

[1]Confidence interval is within clinically acceptable differences

[2]Confidence intervals too wide to reach a conclusion

9. Acceptance Criteria for Positive and Negative Percent Agreements.

For a panel of 100 samples which test positive by the old system, and of which 96 also test positive by the new system (96 out of 100), the lower limit of the 95% two-sided score confidence interval is above 90%. For 30 samples with values close to C_{95}, the 95% two-sided confidence interval for 26/30 (87%) is 70.3% to 94.7%. If, for example, among the 30 samples with low positive concentrations (concentrations close to C_{95} by the old system), only 25 samples test positive by the new system, then the percent of positive results by the new system for the samples close to the cutoff is statistically different from 95% (83% (25/30) with 95% CI: 66.4% to 92.7%). It should be noted that if the percent agreement is 96% or higher at each of three sites, then the lower confidence bound exceeds 90% and a formal calculation of the 95% lower confidence bound may not be needed. The lower bound of the confidence interval for overall agreement is bigger than the smallest of the lower bounds calculated at each site and may be used to demonstrate that the agreement criteria has been met.

10. Calculation of Allowable Total Difference

For each sample of the Comparison panel, calculate the differences between the new system result (Y) and the old system result (X), Y-X (based on CLSI EP21). Also calculate (X+Y)/2. Plot the difference between Y and X, Y-X, against their mean (Y+X)/2 (Bland-Altman plot)[15].

[15] See Bland J.M. and Altman D.G. "Measuring agreement in method comparison studies", *Statistical Methods in Medical Research* 1999; 8; p.135-160.

On the Bland-Altman plot of (Y-X) vs. (Y+X)/2, provide the Allowable Total Difference (ATD) zone around the axis, (Y+X)/2. The ATD zone is established in such a way that 95% of differences between the old system result and the repeated result by the old system fall within the ATD. The ATD zone is expressed as:

$\pm 1.96 \bullet \sqrt{2} \bullet CV \bullet (Y+X)/2 = \pm 2.77 \bullet CV \bullet (Y+X)/2$ for larger values of old system and $\pm 1.96 \bullet \sqrt{2} \bullet SD = \pm 2.77 \bullet SD$ for the low values of old system where CV and SD are the reproducibility characteristics of the old system (see *Establishing SD and percent CV for ATD Based on the Performance of the old system* below). A hypothetical example of the ATD zone on the Bland-Altman plot is provided below (Figure 9):

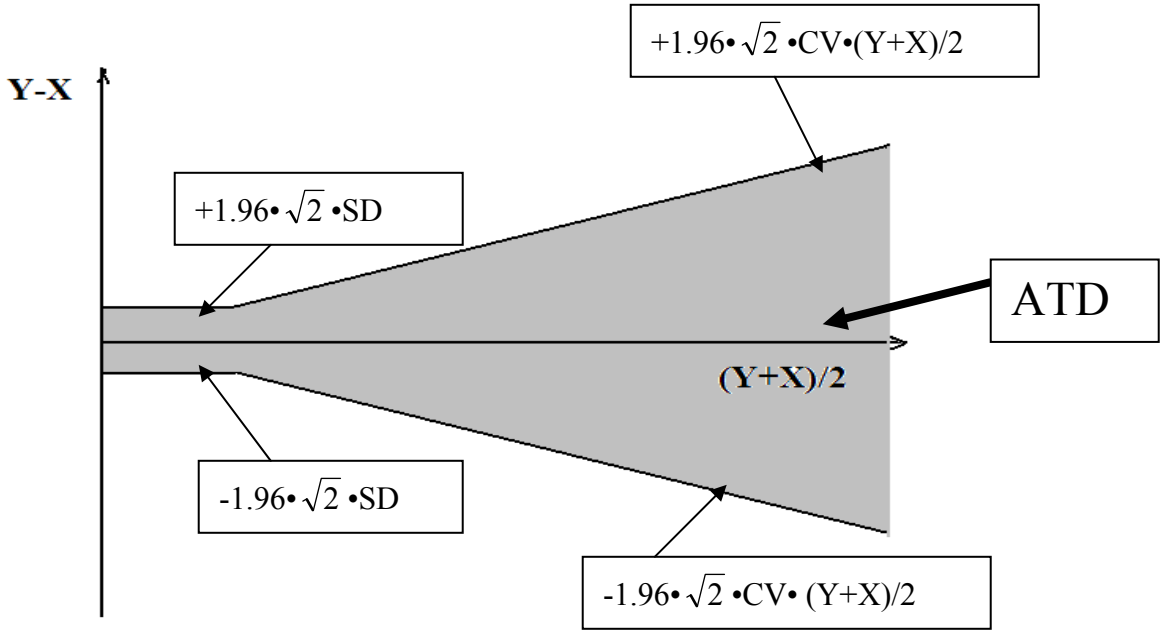

Figure 9. A hypothetical example of the ATD zone on the Bland-Altman plot is provided below.

By the appropriate transformation, a similar ATD zone can be presented on the plane of new system values (Y) against old system values (X), see Figure 7. The expressions of the lines of the ATD zone on the plane Y vs. X are the following:

Lines parallel to the diagonal of the ATD zone:

$$Y = X \pm 1.96 \cdot \sqrt{2} \cdot SD = \mathrm{X} \pm 2.77 \cdot \mathrm{SD}, \quad \text{if } 0 \le \mathrm{Y+X} \le 2 \cdot A$$

For CV≤10% and when each sample in the comparison panel is measured one time by the old system, lines for the "expanding" part of the ATD zone are:

$$Y = X \cdot (1 + \frac{1.96 \cdot \sqrt{2} \cdot CV}{1 - 1.96 \cdot \sqrt{2} \cdot CV / 2}) = X \cdot (1 + \frac{2.77 \cdot CV}{1 - 1.39 \cdot CV})$$

$$Y = X \cdot (1 - \frac{1.96 \cdot \sqrt{2} \cdot CV}{1 + 1.96 \cdot \sqrt{2} \cdot CV / 2}) = X \cdot (1 - \frac{2.77 \cdot CV}{1 + 1.39 \cdot CV}) \quad \text{if } \mathrm{Y+X} > 2 \cdot \mathrm{A}$$

where A is SD/CV.

For CV>0.10 or when each sample is measured more than one time by the old system, an approach based on a ratio of two normal variables is recommended.[16] Examples of ATD lines for CV≥0.10, and when each sample is measured one time by the old system is presented in table below:

	Allowable Total Difference Zone $Line_1 = A_1 * X$ and $Line_2 = A_2 * X$	
CV	A_1	A_2
0.10	0.75	1.32
0.15	0.65	1.54
0.20	0.55	1.81

Establishing SD and percent CV for ATD Based on the Performance of the Old System

For an individual measurement X_i of a given sample by the old system, there is a following expression: $X_i = X_{truei}$ + Mean-Bias + Random-$Bias_i$ +ε_i where deviation of X_i from the true value X_{truei} is composed of a mean bias, a random matrix-related interference component, and a random measurement error[17,18]. Because there are no changes in the assay, it is anticipated that the random matrix- related interferences in both systems are the same. Then the difference between new system and old system measurements of the same sample depends on a random measurement error.

Establishing standard deviation (SD) for the Allowable Total Difference (ATD) zone should be based on consideration of possible variance between the two measurements of the same sample by the old system obtained at different sites. For this, consider the precision profile of the reproducibility study with the between-site component of variance. Another hypothetical precision experiment for the old system can produce a slightly higher SD than in this precision experiment. In order to address this, the observed SD may be multiplied by the appropriate factor (factor $= (1 - 1/(4 \cdot f))^{-1} \cdot \sqrt{f / \chi^2_{5\%}(f)}$, f is a degree of freedom of the estimated SD).

For example, if the degrees of freedom of the SD in the precision study of the old system were 40, then the appropriate factor is 1.236 and the expected maximum observed SD can be as high as 0.148 (=0.120 *1.236). After the appropriate SD or percent CV is established for each concentration in the precision study, the ATD zone can be obtained by smooth interpolation**.**

16 Zhang L., Mathew T., Yang H., K. Krishnamoorthy, Iksung Cho I. (2009): Tolerance Limits for a Ratio of Normal Random Variables, Journal of Biopharmaceutical Statistics, 20:1, 172-184**.**

17 Krouwer JS. Estimating total analytical error and its sources. Techniques to improve method evaluation. *Arch. Pathol. Lab. Med*. 1992; 116:726-731.

18 Linnet K. Boyd JC. Analytical validation of methods – With statistical methods. In: Burtis C, Ashwood ER, Bruns D (eds) *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 4 ed. New York: Saunders, 2006, p.353-407

11. Number of Samples in the Allowable Total Difference Zones

For 150 samples with 95% of the observations (143 /150) falling in the ATD zone, the lower limit of the 95% one-sided confidence interval is above 90%.