Draft Guidance for Industry and FDA Staff:

Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Document issued on: September 7, 2005

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.fda.gov/dockets/ecomments. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this document contact Carol Benson, 240-276-0443 x144 or by email at mailto: carol.benson@fda.hhs.gov

This draft guidance document replaces "Guidance for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Criteria for Waiver", Draft Guidance for Industry and FDA, March 1, 2001.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of In Vitro Diagnostic Device Evaluation and Safety

Preface

Additional Copies:

Additional copies are available from the Internet http://www.fda.gov/cdrh/oivd/guidance/1171.pdf Alternatively, you may either send a fax request to (301) 443-8818 to receive a hard copy of the document, or send an e-mail request to mailto:gene.allen@fda.hhs.gov to request hard or electronic copy. Please use the document number 1171 to identify the guidance you are requesting.

TABLE OF CONTENTS

I. INTRODUCTION 5

The Least Burdensome Approach 7

II. DEMONSTRATING "SIMPLE" 7

III. DEMONSTRATING "INSIGNIFICANT RISK OF AN ERRONEOUS RESULT" – Failure Alerts and Fail-Safe Mechanisms 9

A. Tier 1: Hazard Analysis 10

B. Tier 2: Fail-Safe and Failure Alert Mechanisms 12
1.General recommendations for designing fail-safe and failure alert mechanisms 12
2.External control materials 13
3.Additional points concerning control materials 14
4.Validating fail-safe and failure alert mechanisms, including external control procedures 14

IV. DEMONSTRATING INSIGNIFICANT RISK OF AN ERRONEOUS RESULT – "ACCURACY" 15

A. Clinical Study Sites and Participants 16
1.Clinical testing sites 16
2.Clinical study participants 16
a. Operators 16
(1) Intended operators as study participants 16
(2) Instructions for use 16
(3) Universal precautions 17
(4) Operator questionnaire 17
b. Subjects (Patients) 18
c. Financial disclosure 18
3. Clinical study reports 18

B. Clinical Studies for Tests with Quantitative Results 18
1. Selection of the Comparative Method (CM) 19
2. Clinical Study Design and Statistical Analysis 19
a. Specimen Collection and Sample Preparation 19
b. Statistical Analyses 20
3. Performance Criteria for WM with Quantitative Results 22

C. Clinical Studies for Tests with Qualitative Results 25
1. Selection of the Comparative Method (CM) 25
2. Clinical Study Design 26
a. General Approach for Method Comparison 26
b. Determining Device Performance with Analyte Concentrations near the Cutoff - ... 27
3. Test Performance of Qualitative Tests 28

Statistical notes concerning analyses of percentages 28

V. LABELING FOR WAIVED DEVICES 29

A. Quick Reference Instructions (QRI) 29

B. Quality Control (QC) Labeling Recommendations 29

C. Educational Information 31

VI. SAFEGUARDS FOR WAIVED TESTS 31

VII. REFERENCES 32

Appendix A - SPECIFIC LABELING RECOMMENDATIONS FOR WAIVED DEVICES 34

Appendix B - DEFINITION OF TERMS AS USED IN THIS GUIDANCE DOCUMENT 36

Draft Guidance for Industry and FDA Staff:

Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA): CLIA Waiver Applications

This draft guidance, when finalized, will represent 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

CLIA requires that clinical laboratories obtain a certificate from the Secretary of Health and Human Services before accepting materials derived from the human body for laboratory tests. 42 U.S.C. § 263a(b). Laboratories that perform only tests that are "simple" and that have an "insignificant risk of an erroneous result" may obtain a certificate of waiver. 42 U.S.C. § 263a(c)(2). The Secretary has delegated to FDA the authority to determine whether particular tests (waived tests) are "simple" and have "an insignificant risk of an erroneous result" under CLIA. 69 FR 22849. This guidance document describes recommendations for device manufacturers (you) submitting to FDA an application for determination that a cleared or approved device meets this CLIA standard (CLIA waiver application).

CLIA, 42 U.S.C. § 263 a (d) (3) Examinations and Procedures, as modified by the Food and Drug Administration Modernization Act (FDAMA), reads as follows regarding tests that may be performed by laboratories with a certificate of waiver:

> The examinations and procedures [eligible for certificates of waiver] are laboratory examinations and procedures that have been approved by the Food and Drug Administration for home use, or that, as determined by the Secretary, are simple laboratory examinations and procedures that have an insignificant risk of an erroneous result, including those that - (A) employ methodologies that are so simple and accurate as to render the likelihood of erroneous results by the user negligible, or (B) the Secretary has determined pose no unreasonable risk of harm to the patient if performed incorrectly.

In this document, FDA (we) recommend an approach for you to demonstrate that your device is simple and has an insignificant risk of erroneous result. As part of demonstrating the latter, we recommend studies you can conduct to demonstrate the test is "accurate." This approach is an

alternative to the approach that the Centers for Medicare & Medicaid Services (CMS) and the Centers for Disease Control and Prevention (CDC) published as a Notice of Proposed Rulemaking (NPRM) in the *Federal Register* (60 FR 47534) on September 13, 1995. While we recommend you adopt this approach for waiver applications, you may use another approach that you believe would be appropriate for your device's waiver application if it meets the CLIA statutory requirements.

This draft guidance document recommends you include the following items in your waiver application in order to demonstrate your test is simple and has an insignificant risk of erroneous result:

- A description of your device that demonstrates it is simple to use. (Section II)
- A hazard analysis, including flex studies that identify potential sources of error for your device, and a description of methods you have implemented to mitigate the risk of these errors. (Section III)
- Validation studies to demonstrate the ability of the failure alert and fail-safe mechanisms to mitigate the risk of errors (under conditions of stress). (Section III)
- A description of the design and results of clinical studies you conducted to demonstrate that the device has an insignificant risk of erroneous result in the hands of the intended user (hereinafter operator). (Section IV)
- Proposed labeling with instructions for use that are consistent with a device that is "simple". (Section V)

This document replaces "Guidance for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Criteria for Waiver", Draft Guidance for Industry and FDA, March 1, 2001. Some of the changes compared to the previous document include the following:

- Greater emphasis on scientifically-based flex studies and validation studies, linked to the hazard analysis for each device.
- Recognition that reference methods may not be available for every device type. (However, devices should be traceable to true reference methods of known accuracy, when such methods are available).
- Additional emphasis on use of quality control procedures.
- Greater emphasis on intended users (which may include medical assistants, nurses or doctors and lay people, as appropriate) during studies testing the device.
- Updated study recommendations with emphasis on use of patient specimens, in an intended use environment, over time.

We base the recommendations in this document on our interpretation of the law, our experience with CLIA complexity determinations, and our interactions with stakeholders.
One of the interactions with stakeholders was at an open public workshop on August 14 and 15, 2000. In addition, a proposal presented by AdvaMed (Advanced Medical Technology Association) at the September 2003 Clinical Laboratory Improvement Advisory Committee (CLIAC) meeting, and recommendations proposed by CLIAC during the February 2004 meeting were considered in the development of this guidance.

This document does not address test systems cleared or approved by FDA for over-the-counter or prescription home use, since these automatically qualify for CLIA waiver. 42 USC 263a(b)(3). This guidance document also does not address use of OIVD's (Office of *In Vitro* Diagnostic Device Evaluation and Safety) replacement reagent and instrument family policy[1] for waived devices; that policy does not currently apply to CLIA waiver applications.

We encourage you also, to review the following FDA guidance documents concerning labeling and device design. They are available on the Internet as shown:

- "Write it Right," http://www.fda.gov/cdrh/dsma/897.pdf
- "Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management," http://www.fda.gov/cdrh/humfac/1497.html
- "Guidance on Medical Device Patient Labeling, Final Guidance for Industry and FDA Reviewers" http://www.fda.gov/cdrh/ohip/guidance/1128.html

FDA has also issued a draft guidance entitled "Guidance for Administrative Procedures for CLIA Categorization," www.fda.gov/cdrh/ode/guidance/1143.html. In it, we provide guidance to device manufacturers on FDA's administrative procedures for CLIA categorization.

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.

The Least Burdensome Approach

This draft guidance document reflects our careful review of what we believe are the relevant issues related to CLIA waiver applications and what we believe would be the least burdensome way of addressing these issues. If you have comments on whether there is a less burdensome approach, however, please submit your comments as indicated on the cover of this document.

II. DEMONSTRATING "SIMPLE"

CLIA requires that tests performed by laboratories with a waiver certificate be "simple." We recommend that you determine whether your device is simple, as a first step in the process of deciding whether the device could be a candidate for waiver. You may contact OIVD for input on this issue, prior to conducting clinical studies to support waiver.

Under the approach recommended in this guidance, FDA considers that a simple test should have characteristics such as the following:

- Is a fully automated instrument, or a unitized, or self-contained test.

[1] For a description of this policy, see Guidance for Industry and FDA Staff; Replacement Reagent and Instrument Family Policy, http://www.fda.gov/cdrh/oivd/guidance/950.html

- Uses direct unprocessed specimens, such as capillary blood (fingerstick), venous whole blood, nasal swabs, or urine.
- Needs only basic, non-technique-dependent specimen manipulation, including any for decontamination.
- Needs only basic, non-technique-dependent reagent manipulation, such as "mix reagent A and reagent B".
- Needs no operator intervention during the analysis steps.
- Needs no technical or specialized training with respect to troubleshooting, or interpretation of multiple, or complex error codes.
- Needs no electronic or mechanical maintenance.
- Produces results that require no operator calibration, interpretation, or calculations.
- Produces results that are clear to read, such as 'positive or negative', a direct readout of numerical values, the clear presence or absence of a line, or obvious color gradations.
- Provides instructions and materials for obtaining and shipping specimens for confirmation testing, in cases where such testing is clinically advisable.
- Has test performance comparable to a traceable reference method, as demonstrated by studies in which intended operators[2] perform the test. If a reference method is not available for a test you are proposing for waiver, please contact OIVD to discuss your proposed plan, prior to submitting your application.
- Contains a quick reference instruction sheet that is written at no higher than a 7th grade reading level.

We believe that a test that is simple should not have the following characteristics:

- Sample manipulation is required to perform the assay. (For example, tests that use plasma or serum are not considered simple.) Sample manipulation includes processes such as centrifugation, complex mixing steps, or evaluation of the sample by the operator for conditions such as hemolysis or lipemia.
- Measurement of an analyte could be affected by conditions such as sample turbidity or cell lysis.
- Results need to be reported to a public health department at the state or local level e.g., tests for sexually transmitted diseases, since this is not a requirement that would be explained in the device labeling

[2] In this guidance, intended operator refers to a test operator with limited or no training or hands-on experience in conducting laboratory testing (e.g., medical assistant, nurse, doctor, or an individual with no medical training). Laboratory professional refers to a person who meets the qualifications to perform moderate or high complexity testing, such as a medical technologist (MT) or medical laboratory technician (MLT).

In your waiver application you should describe features of your device that address the issues listed above. Whenever possible, (for example, if your test system consists of a unitized device) you should include sample(s) of the device with your waiver application to aid FDA in its determination of whether it is "simple". You may also schedule a meeting to bring your device to FDA to aid FDA in making this determination.

III. DEMONSTRATING "INSIGNIFICANT RISK OF AN ERRONEOUS RESULT" – Failure Alerts and Fail-Safe Mechanisms

Generally, waived tests should be more robust than non-waived tests. You should demonstrate in your CLIA waiver application that sources of error are controlled or mitigated by fail-safe or failure alert mechanisms. Fail-safe mechanisms are designed with a lock-out function that will ensure that a test system does not provide a result when test conditions are inappropriate or the result is based on faulty test functioning. For example, a system could contain a lockout function that prevents the test from providing a result if the result exceeds the reportable range, if a component malfunctions, or if there is operator error. We recommend that test system design incorporate fail-safe mechanisms whenever possible.

If fail-safe mechanisms are not feasible for all aspects of a system, failure alert mechanisms should be included in the device. Failure alert mechanisms notify the operator of any test system malfunction or problem. Devices with such mechanisms allow the operator to correct the error, or put the operator on notice that the results will be unreliable due to the error. Failure alert mechanisms can include systems such as external controls, internal controls (procedural, or other types) and checks to assure proper functioning of the device electronics. Appropriate fail-safe mechanisms and failure alert mechanisms help assure that a test has "an insignificant risk of an erroneous result".

We recommend a two-tiered approach, outlined below, to demonstrate that your device has appropriate fail-safe and failure alert mechanisms.

Tier 1: Hazard Analysis You should conduct a thorough hazard analysis that identifies all potential sources of error, including test system failures and operator errors. Your hazard analysis should include flex studies, which are studies that stress the operational limits of your test system. This process is fundamental to designing measures that adequately mitigate the sources of error you identify. [See also *Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices* (http://www.fda.gov/cdrh/ode/57.html) for further discussion of hazard analyses)] In your waiver application you should include:

- A report describing the hazard analysis for your device. We recommend that you present this in tabular form and include the hazards and their potential sources of error. (see pp 10-12).
- A summary of the design and results of your flex studies.
- Conclusions you draw from the flex studies.

Tier 2: Fail-Safe and Failure Alert Mechanisms. Once you have identified potential sources of error, you should identify the mitigations, including fail-safe and failure alert mechanisms, that will address these sources of error and conduct validation studies to test the mitigations. In your waiver application you should include the following:

- Identification and physical description of all fail-safe and failure alert mechanisms, including external controls that you recommend for your device, as well as a description of the roles these mechanisms play in mitigating the effect of system failures or user errors you identified in your hazard analysis and flex studies.
- Descriptions of your validation studies and results, tested under appropriate conditions of stress, to support the ability of fail-safe or failure alert mechanisms in your device to prevent or mitigate false results. These studies should support your recommended control procedures and frequencies.
- Instructions for the conditions and frequency of use of external controls, based on validation studies.
- Description of the benefits and limitations of fail-safe and failure alert mechanisms, including all internal and external controls.

A. Tier 1: Hazard Analysis

As noted above, you should identify all potential sources of error by conducting a thorough hazard analysis. This analysis should be part of your risk management process. See the International Standard *ISO 14971:2000, Medical Devices - Application of Risk Management to Medical Devices* for discussion of this process as part of risk management. Based on the results of the hazard analysis, you should conduct flex studies. Flex studies are designed to challenge the system under conditions of stress to identify potential device failures and determine the robustness of the test system. Examples are shown in Table 1 (p. 15)

In your analysis, you should consider multiple skill levels of users, as well as potential instrument and reagent problems. The following websites contain additional information to consider concerning human factors that may affect test performance:
http://www.fda.gov/cdrh/humanfactors/index.html,
http://www.fda.gov/cdrh/humanfactors/resource-manufac.html#2.

Examples of potential sources of error to consider for the hazard analysis and flex studies are listed below. You should consider each of these potential sources of error, as applicable to your device, and also consider any other potential system failures that may be specific to your device.

Operator error/ Human factors

- Use of incorrect specimen type.
- Incorrect application of the specimen on the device.
- Incorrect placement of device (e.g., non-level surface).

- Incorrect placement of reagents including strips, or other components that contain reagents.
- Use of incorrect reagents, for example, reagents that are not specific for the particular device or lot, or generic reagents.
- Incorrect order of reagent application.
- Use of incorrect amount of reagent.
- Incorrect timing of analysis (e.g., specimen application, running the test, or reading results).
- Incorrect reading of test results.

Specimen integrity and handling

- Error in specimen collection.
- Use of inappropriate anticoagulant.
- Clotted specimen.
- Error in specimen processing and handling.
- Incorrect specimen transport and/or storage.
- Presence of interfering substances.
- Presence of bubbles in the specimen.

Reagent integrity (Reagent viability)

- Use of improperly stored reagents.
- Use of outdated reagents.
- Use of improperly mixed reagents.
- Use of contaminated reagents.

Hardware, Software and electronics integrity

- Power failure.
- Repeated plugging and unplugging of the device.
- Hardware failure.
- Software failure, see *Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices* (http://www.fda.gov/cdrh/ode/57.html)).
- Electronic failure.

- Physical trauma to unit.

Stability of calibration and internal controls

- Factors that affect calibrator and calibration stability, including determination of calibration stability over time and after power failures.
- Factors that may interfere with calibration.

Environmental factors

- Impact of key environmental factors (heat, humidity, sunlight, surface angle, device movement, etc.) on reagents, specimens, and test results.
- Impact of key environmental factors (including electrical or electromagnetic interference) on instruments, if appropriate.

B. Tier 2: Fail-Safe and Failure Alert Mechanisms

1. General recommendations for designing fail-safe and failure alert mechanisms

You should consider incorporating fail-safe mechanisms, as a first preference. When designing fail-safe or failure alerts mechanisms, consider including the following, where appropriate.

- Lock-out functions that do not allow output of results if controls or system checks are not completed, or if the test does not give expected results.
- Lock-out functions that do not allow output of results if the device was mishandled (e.g., dropped).
- Monitors of environmental conditions (e.g., indicator desiccants) incorporated into the test system, or the kit container, to alert the user to environmental conditions that are outside of the recommended storage conditions.
- Battery checks.
- Internal procedural controls to flag procedural problems such as improper sample flow, incorrect use of components, or improper addition of specimen.
- Internal non-procedural controls, (e.g., for checking the integrity of the reagents).
- Controls to check that electronic features of the device are within specifications.
- External control material.

When designing controls, you should consider the unique features of the test system and link control procedures to the robustness of the assay, as determined by your flex studies.

The controls you devise to mitigate the risks you identify may be based on standard laboratory control procedures typical for laboratory-based methodologies (e.g., testing external materials, at two levels, once per shift or on each day of testing), or may be a combination of features, such as those listed above, that ensure complete system quality monitoring. When designing packaging for your device, you should also consider that the number of tests per kit should depend on stability of reagents or robustness demonstrated through testing.

When appropriate, you should incorporate capabilities into the test system software that allow for data retention, identification of outliers, and trend detection, in order to alert the user to the occurrence of random or systematic errors.

Procedural controls which are typically internal, are desirable for waived devices. However, these types of controls generally do not replace external controls, especially because they often only control for adequate volume. Your flex studies and validation studies should evaluate the sensitivity of internal control reagents to all applicable test system errors. The total QC system (including all control procedures and internal checks) should control for all aspects of test performance, including electronic aspects and integrity of reagents.

You should include aspects of the device design that are controlled and maintained by the manufacturer as mitigations. We do not recommend you identify training as a sole means of mitigating potential sources of harm.

2. External control materials

Whenever feasible, you should include external control materials in the test kit.

You should alert operators about control procedures and the availability of control materials and integrate instructions for external control testing within the test procedure instructions (package insert and Quick Reference Instructions), in order to increase the likelihood that operators will perform QC correctly. In the test instructions, you should specify minimum frequency for running controls, and include recommended levels of control materials that correspond to medical decision levels. The labeling should indicate in bold why external controls are important and the repercussions and consequences of not performing all QC procedures.

In addition, when control materials are not included in the test kit, you should also recommend, in the package insert and Quick Reference Instructions the use of specific control material(s) that will ensure optimal verification of the test system performance. Providing or recommending external control materials may not be critical in those limited cases where sufficient fail-safe mechanisms are in place for the entire system. We are unaware of such systems currently, though some may be developed in the future. In such cases, you should explain your rationale for omitting these control materials, in your waiver application.

External control materials for waived tests should be ready to use, or employ only very simple preparation steps, e.g., breaking a vial in order to mix liquid and dry components of the control material. For both quantitative and qualitative tests, the levels of the control materials should correspond to the medical decision level(s) relevant to the indications for use for your test. More than one level may be needed in order to ensure accuracy for quantitative tests. The control material should be traceable to a reference material whenever possible.

You should describe, in your application, how you established the limits of acceptable performance for the control material that you include with, or recommend for use with, your device; and how you demonstrated that use of these limits provides an adequate assessment of whether the test is performing properly. For quantitative tests, you should consider the precision of the test system, as well as the total allowable error for the particular analyte, when setting external quality control limits of acceptable performance. Ranges that are too broad may be incapable of reliably detecting unacceptable levels of imprecision or bias.

Control materials should mimic patient samples as closely as possible. When the matrix of the material differs from that of the specimen, you should determine and describe in your application how these differences may affect or limit the information provided by the control result. You can accomplish this by testing control materials in parallel with actual patient samples of similar known values and comparing the results of the control material and patient samples with respect to precision or bias observed. You should account for matrix effects when setting the limits for control material to be used with your test.

3. Additional points concerning control materials

If you did not previously submit information addressing the items below in your premarket submission, you should provide them in your waiver application:

- Opened and unopened control material stability data

  You should include stability data and acceptance criteria for opened and unopened control material. The term "unopened" refers to shelf-life stability whereas "opened" refers to reconstituted or opened conditions. You should support stability claims with accelerated studies, ongoing real time studies, or real time data.

- Lot-to-lot reproducibility, conducted on at least three consecutive lots of control material.

4. Validating fail-safe and failure alert mechanisms, including external control procedures

You should conduct studies that validate all fail-safe and failure alert mechanisms (including any procedures you recommend that use external control materials) that will address all the causes of test errors that you identified in your hazard analysis. These studies should be conducted under conditions that stress the device in order to

demonstrate how fail-safe and failure alert mechanisms respond to such conditions. You should describe your validation studies and results in your waiver application, and indicate how the results support the ability of fail-safe and failure alert mechanisms to detect and mitigate test errors. You should include a description of how your recommendations for external control materials and procedures (including frequency) are supported by your validation studies, and confirmed by the clinical studies described in Section IV, below.

Table 1 - Examples of approaches to flex studies and control validation studies under conditions of stress

POTENTIAL SOURCE OF ERROR	EXAMPLES OF FLEX STUDIES	EXAMPLE OF VALIDATION STUDIES
Operational storage is 2-4° C. What happens when the kit is stored improperly?	Environmental studies included storing the kit at 0°, 2°, 10°, 25°, and 37° C. Studies showed that when frozen, or stored at 25° C for over 3 days, the device failed.	Studies to validate that fail-safe mechanisms, control procedures, or failure alerts alert the operator to frozen conditions, or storage at 25° C for more than 3 days.
Procedure is to add 3 drops. What happens when an improper number of drops are added to the test procedure?	Flex studies consist of adding 1, 2, 3, 4, 5, and 6 drops and observing when incorrect results are obtained. Studies show that <2 drops or >5 drops give erroneous results.	Studies to validate that fail-safe mechanisms, or failure alerts, including external control procedures, alert the operator of an error when <2 drops or > 5 drops are added.

IV. DEMONSTRATING INSIGNIFICANT RISK OF AN ERRONEOUS RESULT – "ACCURACY"

In this guidance document we use the term "accurate" tests to refer to those tests that are comparable to a traceable method, in which the results of measurements can be related to stated references. (Also see reference [1].)[3] To demonstrate that your device is "accurate" in the hands of the intended operator, we recommend that you perform prospective clinical studies of the device proposed for waiver, using patient samples[4] collected in the intended testing environment. In this way, the studies will demonstrate, as closely as possible, how the device performs on actual clinical specimens by intended operators under the conditions of intended use.

In this document, we describe the study designs we recommend for a quantitative test and a qualitative test. A quantitative test is a test that gives results expressing a numerical amount or level of an analyte in a specimen; a qualitative test is a test that provides only two responses (i.e.,

[3] Literature references listed in Section VII are indicated by numbers in brackets.
[4] Spiked samples may also be appropriate for a portion of the study; See Section B2a, below.

positive/negative or yes/no) [2]. If your test does not fit the paradigm of a quantitative, numerical test or a qualitative, two-response test, please consult with OIVD.

You should evaluate test performance in a setting designed to replicate, as closely as possible, the actual intended clinical use setting. Therefore, the study design should include:

- intended clinical testing sites,
- intended operators as study participants,
- intended sample type and matrix whenever possible; and
- testing over time, as in typical intended use setting.

A. Clinical Study Sites and Participants

1. Clinical testing sites

You should conduct the clinical study to support CLIA waiver at a minimum of three intended use sites at different demographic locations (e.g., out-patient clinic, physician's office) representative of the type of site(s) where the device will be used. In your waiver application, you should present a brief description of each site, along with its name and address.

2. Clinical study participants

a. Operators

(1) Intended operators as study participants

You should enroll individuals who represent anticipated operators of the device you propose for CLIA waiver. We recommend that you record and tabulate the information listed below concerning each participant to demonstrate that participants meet the definition of intended operators, and include this in your CLIA waiver application.

- age
- gender
- education (including experience and training)
- occupation

In addition, for each study site, we ask you to report the same information (i.e., age, gender, education, and occupation) on all potential intended operators at the testing site.

(2) Instructions for use

You should provide the intended operators who participate in the study with only the proposed package insert and/or Quick Reference Instructions. Study participants should receive no training, coaching, prompting, or written or verbal instructions beyond the written test procedure. They should have no opportunity to discuss the test with, or otherwise coach, or observe each other. You should include, in your waiver application, the instructions you provided to test operators participating in the study. The instructions should be those you plan to provide to

operators when the test is marketed, and should include all control procedures you plan to recommend to device users, after it is marketed .

(3) Universal precautions

You are required to conduct clinical studies under conditions that comply with Occupational Health and Safety Administration (OSHA) regulations pertaining to biological hazards ("universal precautions"), 29 CFR 1910.1030.

(4) Operator questionnaire

You should develop an operator questionnaire to be filled out by all test operators participating in the study. This questionnaire should be designed to help assess whether the participants understood how to use the device correctly. It is important that the questionnaire be given to test operators *after* the completion of the clinical study, so that the questions do not bias the participants during the study. Examples of questionnaire items are listed below.

Some questions may ask operators to indicate agreement on a 1-5 scale (1=strongly disagree; 5=strongly agree). For example,

- The instructions are easy to understand.
- The instructions are difficult to follow.
- It is easy to apply the sample.
- It is easy to see and understand the test results [appearance of the line, change of color etc…]
- The control line was always clearly present and easy to read.
- The instructions clearly explain what to do if a test result does not appear, or is invalid.
- I tested control materials according to the package insert.
- If I did not have any test instructions, I would perform the test without instructions.
- I needed help from someone the first time I did the test.

True/false questions may enable you to determine whether operators understand the meaning of test results. For example:

- This test is a screening test.
- This is a confirmatory test
- Positive results should be confirmed.
- An invalid test means the result is negative.
- Testing control materials is not necessary for this test.

You should also strongly encourage general comments by the operators. We recommend that you include your survey questions and results with your CLIA application.

b. Subjects (Patients)

You should ensure that subjects from whom you will obtain specimens for the clinical study meet inclusion and exclusion criteria corresponding to the intended use population of the test. Once a subject has met appropriate inclusion and exclusion criteria, he/she should be informed of the study and invited to participate in the CLIA waiver study. We recommend you obtain informed consent for each subject in the CLIA waiver study.

c. Financial disclosure

If clinical investigators are involved in the clinical study, a Financial Disclosure Statement may be required. For advice on whether the financial disclosure rule applies, please refer to the CDRH guidance, "Guidance for Industry: Financial Disclosure by Clinical Investigators," http://www.fda.gov/oc/guidance/financialdis.html or the 21 CFR Part 54, Financial Disclosure by Clinical Investigators.

3. Clinical study reports

You should report results of the clinical study to support CLIA waiver by each intended site, and overall if appropriate. Reports should include the following:

- protocol description,
- number of subjects (i.e., patients) studied,
- procedures for subject inclusion and exclusion,
- description of the subject population,
- description of how specimens were collected and stored,
- masking techniques,
- discontinuations,
- complaints, device failures and replacements,
- any invalid results and how these were handled,
- information about QC procedures that were performed by intended users,
- pertinent tabulations,
- annotated line listings of results (including electronic versions), and
- clear descriptions and presentations of the statistical analyses.

You should not remove "outliers". In the event that a part of the collected data is not included in the analyses, you should clearly identify and provide justification for removing these data.

B. Clinical Studies for Tests with Quantitative Results

The clinical studies to support CLIA waiver should compare results obtained with the device proposed for CLIA Waiver (WM) to results obtained by a Comparative Method (CM). The CM for the clinical studies for quantitative tests should be performed in a laboratory setting by laboratory professionals.

1. Selection of the Comparative Method (CM)

The following Comparative Methods are presented in the order in which we recommend you select them, if available. Thus, we recommend you use a CM of type A, if available, and if not, that you select type B, and provide adequate justification for selection of this method in your waiver application. *If CM of types A or B are unavailable, we strongly recommend that you consult with OIVD to discuss potential comparison with a CM of type C or other well-documented method*.

- **CM of type A:** This is a reference method (RM) which has been thoroughly investigated. It has been shown to yield values having trueness and precision of measurement such that the RM can be used to assess the trueness of other methods for the same quantity or for assigning values to reference materials (for details, see Harmonized Terminology Database at the CLSI website: www.clsi.org).
- **CM of type B**: This is a traceable method in which the results of measurement can be related to a stated RM, usually a national or international standard, through an unbroken chain of calibrations of a measuring system or comparisons where measurement uncertainties have been documented at every step in the procedure. For details on traceability, see [1]. You should provide a description of the traceability chain (the mathematical relationships between the measurement results with stated uncertainties) and how the traceability chain was established (calibration transfer protocols). For instance, if the mathematical relationships were described by linear regression relationships, then, for traceability of the CM of type B to be achieved, observed values of unit slopes and zero intercepts are expected with acceptable levels of uncertainty. The CM of type B provides measurement values with the same degree of trueness as the RM and/or reference materials. The values of the CM of type B are close to the values of RM.
- **CM of type C**: This is a traceable method where a deviation from unit slope or a deviation from a zero intercept (relative to the RM) in the linear regression relationships of the traceability chain may be clinically tolerable (see [1]). In this case, tolerance limits (distinct from uncertainty measurements) will depend on the state of development of methods of measurement and the medical uses to which the results of the CM are to be applied. You should provide equations that describe the relationship of the CM to the RM and cite the relevant sources for these equations. As noted earlier, we strongly recommend that you contact OIVD prior to developing a CLIA waiver submission using a CM of type C.

2. Clinical Study Design and Statistical Analysis

a. Specimen Collection and Sample Preparation

You should conduct the clinical study to support CLIA waiver with a minimum of 360 patient samples from the intended use population. The samples should be obtained at a minimum of 3 clinical sites that are representative of both the intended patient population and the intended operators. Between 1 and 3 intended operators per site should be selected; however, a minimum of 9 operators should participate. For example, one could use 6 sites with 1 or 2 operators per site to reach the

minimum of 9 users. The patient samples should be as equally distributed among the operators as possible. The samples should be from consecutive patients over an appropriate period of time (this period may depend on the specific clinical site, the prevalence of disease, or other factors). We suggest a 1 month period may be useful, and recommend not less than 2 weeks. The goal is to assess how well the device you propose for CLIA waiver works in practice, recognizing that operators may have other duties.

You should ensure that the 360 patient samples span the measuring range of the device and adequately represent all possible values of the CM test [3,4]. When samples are divided into low, medium and high medically relevant intervals (according to the CM values) with approximately the same number of patients in each, 120 observations are recommended to estimate the total analytical error within each of these intervals [3,5]. If medical decision points occur close to the boundary of two of these intervals, we recommend that you change the boundary to be farther away from the medical decision point[4].

We believe that actual patient specimens provide the best assessment of a device proposed for CLIA waiver in the hands of intended operators. However, in some cases this might be impractical, or samples might be distributed insufficiently to challenge the measuring range, or generate values near medical decision points. In such cases, you should substitute, or supplement, up to 60 actual patient specimens with spiked or otherwise contrived matrix-specific specimens, consistent with the intended use of the device. A suggestion concerning the distribution of CM values in your data may be found for some analytes in Table 1 of [4]. You should describe how you prepare the contrived specimens and validate the assigned values. It may also be appropriate to use archival patient samples. *If you believe that more than 60 of the samples should be spiked, contrived or archival, please contact OIVD to discuss*.

Each sample should be split in two parts: one part should be tested by the intended operator as a study participant using the WM, and the other part should be tested by a laboratory professional using the CM. If the sample cannot be split into parts, then a second sample from the same patient should be obtained. *If the order in which the samples are collected impacts the results of testing, we strongly encourage you to consult OIVD concerning your clinical study design.*

b. Statistical Analyses

You should compare the WM results with CM results by an appropriate regression analysis and calculate the Total Analytical Error (based on analyses of differences) [3]. You should provide the following information based on each site separately, as well as a combined analysis over all sites:

(1) Descriptive statistics

- Descriptive statistics for both the WM and CM results, including the number of results, mean, standard deviation (SD), minimum, median, and maximum, and side by side box-and-whiskers plots for both WM and CM values [6].

- Scatter plot of the results, where the WM results are on the Y-axis and the CM results are on the X-axis. 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.

- If the CM is of type C, then, additionally, you should provide the scatter plot of the results with WM results on the Y-axis, and the calculated RM results on the X-axis. RM is calculated using the CM results and mathematical relationships, as described previously under CM of type C section, in the comparative method section.

(2) Regression Analysis

- Results of an appropriate regression analysis. The regression method you choose should account for the random errors associated with the WM and CM. You should present results for the combined data and then separately for each site. A Deming regression procedure or one of several similar methods may be appropriate [7], [8] and [9]. Since the methods may vary in their assumptions, you should include some justification of the choice of procedure in the application. If the random error of the CM is negligible compared to the random error of the WM, and the standard deviation of the random error of the WM is roughly constant over the range of CM values, then ordinary least-square regression analysis is also appropriate [6].

- The 95% confidence intervals of the slope and intercept from the regression in a., above. We recommend that for the data from each site and for combined data, you draw the regression line on the corresponding scatter plots and plot the fitted lines (for each site and overall) on the same corresponding figures. Using the regression equation, you should calculate the systematic bias at medically important points. (For details, see section 6.1 in [4])

- For the CM of type C, you should include the appropriate regression analysis between WM results and calculated RM results, in a similar way to b.

Please note that correlation alone is often inappropriately used to assess agreement, and we caution you against such an approach.

(3) Total Analytical Error

For purposes of this guidance, we use the term "total analytical error" as an interval that contains a specified proportion (e.g., 95%) of the distribution of differences between the WM and CM values (for the CM of types A, B and C) as well as the difference between WM and RM values (for the CM of type C). Total analytical error may also be expressed in terms of relative differences. Relative differences are defined as: the WM result minus the CM result, divided

by the CM result. For details on the total analytical error, see [3]. We recommend that you provide the following information concerning total analytical error:

(a) For every sample tested by WM and CM, you should calculate a difference, (WM result minus CM result). For the CM of type C, you should also calculate the difference between the WM result and the calculated RM result (WM result minus RM result). In a similar way, for every sample, you should also calculate a relative difference: WM result minus CM result, divided by CM result. (In addition, with a CM of type C, you should calculate WM result minus RM result divided by RM result).

(b) You should provide the Bland-Altman plot with the CM values on the X-axis and differences between WM and CM values on the Y-axis. You should also provide another plot with CM on the X-axis and *relative* differences on the Y-axis. For the CM of type C, you should also provide the Bland-Altman plot, with RM values on the X-axis and differences (including, relative differences) between WM and RM values on the Y-axis. For details see [3] and [10].

(c) You should divide the measuring range into three medically relevant intervals (low, medium, and high) as described earlier, where each interval contains approximately the same number of points (about 120). The medically important points should not be at the boundaries of these intervals. For details, see [4]. These ranges will be used for further analyses (step d., below).

(d) For each interval (low, medium, and high), you should calculate total analytical errors of differences or relative differences (whichever is appropriate) for 95% and 99% of differences. Usually, calculation of relative differences is more appropriate for intervals containing high values and calculation of differences is more appropriate for low intervals. You should also provide a histogram of the relative differences (percentages), or actual differences, for each interval. You should calculate both the mean and standard deviation of these differences. You should identify the 2.5^{th} and 97.5^{th} percentiles used in the calculation of the total error for 95% differences and 0.5^{th} and 99.5^{th} percentiles used in the calculation of the total error for 99% differences.

3. Performance Criteria for WM with Quantitative Results

You should establish performance criteria for your device *before* you begin the clinical study, in order to objectively evaluate the WM. You should establish criteria for the following zones and demonstrate that your device meets these criteria:

- The Allowable Total Error (ATE) zones.

The ATE zones are established for 95% of differences between the WM and CM

values (for the CM of types A, B, and C), or between the WM and RM values (for the CM of type C). Values of WM that fall within the ATE zones are values that can be tolerated without invalidating the medical usefulness of the WM results.

- The Zones of Limits for Erroneous Results (LER).

Values outside the LER zone are considered dangerous; when WM values fall outside the LER zones, potential harm can occur to the patients if these results are utilized in medical decision-making. Therefore, your device should not have WM results in the LER zones.

Usually, ATE and LER are expressed as percents of CM values for higher values, and as a flat limit for low values.

Examples of the ATE and LER zones are the corresponding zones in the Consensus Error Grid [11] and Clarke Error Grid [12] plots for glucose. The Clarke Error Grid is illustrated below.

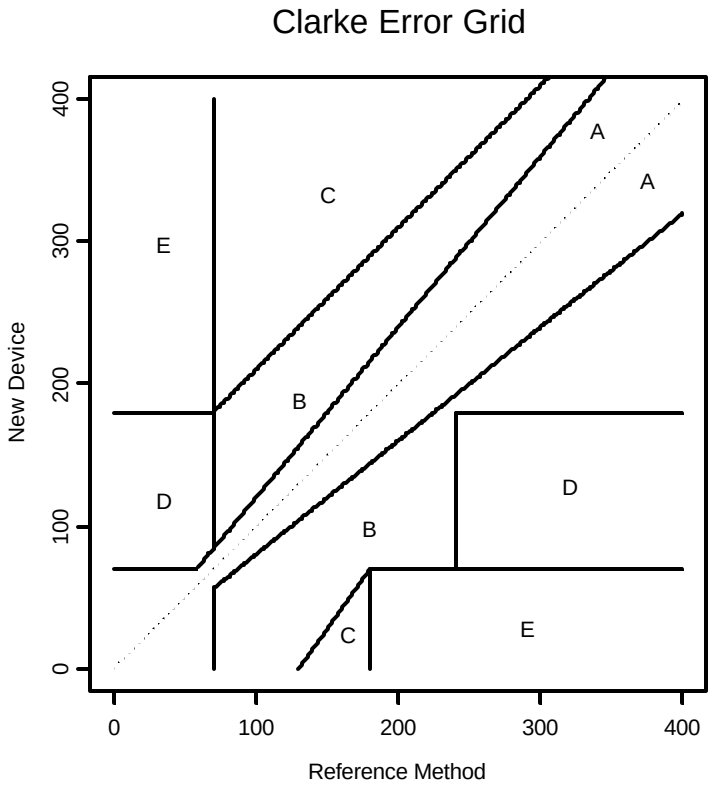

Figure 1. Clarke error grid.

Note that these plots are for glucose only and an Error Grid would be unique for each analyte. In this Clarke Error Grid Analysis, zone A is the zone for the ATE and zones D and E are zones outside the LER. For this analysis, zone A is regarded as clinically accurate if WM is below 70 mg/dL, when CM is below 70mg/dL, and if WM falls

within 20% of CM, when CM≥70mg/dL. Zones D and E correspond to serious discrepancies between CM and WM resulting in misdiagnosis or erroneous treatment of the patient.
Several sources of information can be used to establish criteria for ATE and LER zones for the analyte your device measures. OIVD *will review your criteria for ATE and LER and determine whether they are set at a sufficiently low value to meet the statutory requirement of rendering the likelihood of erroneous results by the user negligible.*

For analytes that have existing performance limits for professional use (e.g., those listed in the CLIA regulations (see 42 CFR 493.929, http://www.phppo.cdc.gov/CLIA/regs/subpart_i.aspx#493.929)), these limits should be used for ATE in the following way. Let a value %R be the allowable percent of deviation from the target value described in the CLIA regulations cited above. For example, the value of %R equals 7% for hemoglobin and 25% for theophylline. In the cases where the allowable percent of deviation is higher than 20%, the value of %R should be adjusted downwards to 20%. In the example above, the value of %R would be 7% for hemoglobin and 20% for theophylline. Once the value of %R is defined, the zone of ATE is bounded by CM ± %R * CM. If, when values of the CM and WM are both low, one can tolerate a larger value of %R * CM without invalidating the medical usefulness of the WM results, then the zone of ATE for these low values can be established as CM ± D, where D is a fixed number. Thus, the ATE zone can be defined as CM ± D for low CM values (i.e., below a defined threshold value) and CM ± %R * CM, for high CM values that exceed the threshold. For example, for the analyte lithium, the ATE zone can be defined as CM ± 0.3 nmol/L, when CM < 1.5 nmol/L and CM ± 20% * CM when CM = 1.5 nmol/L.

For analytes not listed in the CLIA regulations at subpart I, other criteria may be acceptable but you should consult with OIVD. For example, the ATE and LER zones could be based on medical decision-making, consideration of intra-individual variations, needs for accuracy of the samples within the reference intervals, or other approaches, as applicable to the particular analyte. You should provide the literature and appropriate documentation to support the ATE and LER you establish. WM's that have a high random bias, or a systematic bias, may have a greater challenge meeting acceptable criteria, and may not be suitable for CLIA waiver.

We recommend that you include the ATE and LER zones on the scatter plots in your statistical analyses (WM vs. CM for CM of types A, B or C and also WM vs. RM for CM of type C).

For the zone of the ATE, you should report:

- The percentage of the observations that fall within the ATE zone for the low, medium, and high ranges of the CM.
- The percentage of the observations over the entire measuring range that fall within the ATE zone, with a 95% exact lower confidence bound (see Statistical Notes in Section C3, below, for more information on exact

confidence bounds). This lower bound should exceed 93%. This is because for a sample consisting of 360 observations, with 95% of observations (342 out of 360) falling in the ATE zone, the lower one-sided 95% exact confidence bound is 93%.

For the zone of the LER, you should report

- The percentage of the observations that fall within the zone for the low, medium, and high ranges of the CM.
- The percentage of the observations over the entire measuring range that fall within the zone, with a 95% lower confidence bound. This lower bound should exceed 99%. This is because for a sample consisting of 360 observations and all observations (360 out of 360) falling in the LER zone, the lower one-sided 95% exact confidence bound is 99.2%.

C. Clinical Studies for Tests with Qualitative Results

There are two parts to the clinical study design for qualitative data. The first part (discussed in Section C2a, below) is a method comparison and is similar to the study design discussed above for quantitative results. The objective in this part is to compare CM and WM in terms of performance. The second part (discussed in Section C2b, below) is designed to identify problems that may occur when the analyte is near the cutoff concentration which differentiates a positive from a negative outcome. As in the case of a test with a quantitative outcome, the CM for the CLIA waiver clinical studies for qualitative tests should be performed in a laboratory setting by a laboratory professional.

1. Selection of the Comparative Method (CM)

The following CMs are presented in the order in which we recommend you select them, if available. Thus, we recommend you use CM of type A, if available and, if not, that you use CM of type B. If neither CM of type A nor CM of type B is available, then we recommend you use type C. Similarly, for CMs of types D and type E. If you choose a comparative method other than type A, you should provide justification for this in your application. *If you are considering use of a CM of types C, D or E, or another well-documented method, we strongly recommend that you contact OIVD prior to conducting the clinical study in support of a CLIA Waiver application.*

- **CM of type A**: A quantitative reference method such as those outlined as CM of type A under the section on quantitative tests with the appropriate cutoff value for the positive and negative results.
- **CM of type B**: A quantitative traceable method such as that outlined as CM of type B under the section on quantitative tests with the appropriate cutoff value for the positive and negative results.
- **CM of type C**: A quantitative traceable method, such as that outlined as CM of type C under the section on quantitative tests, with the appropriate cutoff value for the positive and negative results.
- **CM of type D**: A qualitative reference method (for examples of the qualitative reference methods, consult with OIVD).

- **CM of type E**: A qualitative method which was tested by reference specimen panels (e.g., panels of samples prepared by well recognized institutions, such as WHO, CDC, NIST). W*e recommend that you contact OIVD for input concerning your choice of specimen panels.*

2. Clinical Study Design

a. General Approach for Method Comparison

As in the study for tests with quantitative results, the samples should be obtained at a minimum of 3 intended use sites that are representative of both the intended patient population and the intended operators. Between 1 and 3 intended operators per site should be selected, however a minimum of 9 operators should participate. Thus, one could use 6 sites with 1 or 2 operators per site to reach the minimum of 9 operators. The patient samples should be approximately equally distributed among the operators. The samples should be from consecutive patients over an appropriate period of time. (This period may depend on the specific clinical site, the prevalence of disease, or other factors). The goal is to assess how well the device proposed for CLIA waiver works in practice, recognizing that operators may have other duties. A one month period for the clinical study to support CLIA waiver may be useful, especially when the prevalence rate is low. The goal should be to have a minimum of 120 samples positive by CM and a minimum of 120 samples negative by CM. However, the overall sample size may well exceed 240, since one cannot anticipate the CM outcome for each patient's data at the time of sample collection. Each operator should observe a minimum of 5 positive and 5 negative samples. However, we stress that the samples should be masked with respect to the CM outcomes. Some clinical sites may differ in their overall prevalence rates. For these situations, alternative study designs and corresponding statistical analyses can be considered. Please contact OIVD for advice prior to conducting a study based on an alternative design.

Tests for low prevalence diseases can present challenges in obtaining at least 120 samples positive by CM. Some approaches are to expand the duration of the CLIA waiver accuracy study, or to include a site with higher disease prevalence among the study subjects. In some cases, the seasonal variations in the prevalence of disease should be taken into account for the study design. If the prevalence of the disease is low, the population can be enriched by archival patient samples (if applicable). We believe that actual patient specimens provide the best assessment of a device in the hands of intended operators. However, in some situations, when disease has low prevalence, you may substitute or supplement actual patient samples with spiked or contrived matrix-specific samples, consistent with the intended use of the device. In such cases, no more than 10% of the samples should be spiked, artificially constructed or archival. *If you believe your device calls for a study design that includes more than 10% contrived or archival samples, you should consult with OIVD for recommendations prior to your study.*

Each sample should be split in two parts: one part should be tested by the intended operator, employing the WM; the other part should be tested by a laboratory

professional employing the CM. If the sample cannot be split into parts then a second sample from the same patient should be obtained. *If the order in which the samples are collected impacts the results of testing, we strongly encourage you to consult OIVD concerning your study design.*

Statistical analysis of method comparison results for qualitative tests:

For the positive and negative patient samples, as determined by CM, you should calculate positive and negative agreement estimates along with the lower bound of the 95% two-sided confidence interval. You should perform calculations for each site, individually and, if appropriate, for all sites combined. Exact confidence intervals are the recommended method for this purpose. See [13-15] and Statistical notes concerning analysis of percentages, in Section IV C 3 below.

b. Determining Device Performance with Analyte Concentrations near the Cutoff

We recommend the following study design for determining performance with samples containing analyte concentrations near the clinical cutoff. You should select 3 intended use sites for this part of the study.

1. Prepare 60 aliquots of one sample with a concentration at which the professional operators of the CM obtained positive results 95%-99% of the time. We refer to such samples as weak positive samples in the discussion below. In your waiver application, you should provide information on how you prepared the samples at this concentration. In some cases, it may be necessary to pool patient samples, or to dilute a sample to create a sufficient amount of material for 60 aliquots.

2. Prepare 60 aliquots of a weak negative sample. We define this as a sample with an analyte concentration at which the professional operators of the CM obtained negative results 95%-99% of the time. For some tests when the cutoff value is set in such a way that only samples with no analyte produce negative results 95%-99% of the time, the prepared aliquots should not have any analyte. As with the weak positive samples, you may pool patient samples or dilute a sample to create enough material for 60 aliquots. For additional discussions on the definitions of weak positive and negative, see [2].

3. These 60 weak positive aliquots and 60 weak negative aliquots should be approximately equally distributed across all operators with 20 samples of each type at each of the three participating sites. You should ensure that the labels on the aliquots are masked as to their true designation, and that each aliquot has a unique label. You should also ensure that samples are tested in <u>random order</u> during the study.

Statistical Analysis for Results of Analyte Concentrations near the Cutoff:

1. For each of the two categories of aliquots (weak positive and weak negative), you should calculate the percent of positive results for the weak positive sample and

the percent of negative results for the weak negative sample, with their 95% two-sided confidence intervals. In the waiver application, you should provide these results for each site, and overall.

2. We recommend that you compare the percents of positive results for the weak positive sample among the three sites (20 measurements per site) by using a Fisher-Freeman-Halton test (a generalization of Fisher's exact test, see [16]). You should conduct a similar test for the weak negative results.

3. Test Performance of Qualitative Tests

We recommend you perform statistical analyses such as the following to demonstrate the "accuracy" of your device.

For the first parts of the study design, described in section c2a, above:
The observed positive and negative agreements between the test proposed for waiver and the CM should be 95% or greater; but in all cases the exact lower confidence bound for the two sided 95% confidence interval should be equal to or greater than 91%. See [13-15] and Statistical notes, concerning analyses of percentages, below. *In some cases, a higher percent agreement and a higher value for the lower confidence bound may be needed to reasonably assure that the WM is "accurate".*

For the second parts of the study design, described in section c2b, above:

- The percent of positive results for the 60 aliquots of the prepared, weak positive samples should be close to 95%. That is, approximately 57 out of 60 of these samples should yield positive results.
- The percent negative results for the prepared weak negative samples should also be approximately 95%.
- The differences in percents of the positive results among the three sites for the weak positive sample and differences in the percents of the negative results among the three sites for the weak negative sample should not be clinically or statistically significant at α=0.05.

Statistical notes concerning analyses of percentages

The following are additional recommendations for performing statistical analyses:

Exact confidence limits using the Clopper-Pearson method for percent positive agreement and percent negative agreement can be calculated in many software packages, or can be obtained from published tables. You may use this approach. However, exact methods tend to produce intervals that may be too wide in some cases. As an alternative, Agresti and Coull [16] recommend a calculation called the score confidence interval. This is also described in [2]. Either the score method, or the exact method for calculating confidence intervals would be appropriate. One-sided confidence bounds (i.e., those used to estimate a lower bound) are suggested in evaluating the performance of the ATE and LER for quantitative tests. We recommend that exact methods such as those based on the Clopper-Pearson method also be used; these are also available in many software

packages.

We expect both the true negative percent agreement and true positive percent agreement to be above the lower confidence bounds in 95% of the waiver studies. Thus we suggest that you calculate two-sided 95% confidence intervals for positive percent agreement and negative percent agreement. We recognize that this is easier to do than two one-sided confidence bounds, one for percent positive agreement and one for negative percent agreement. If you wish to calculate two one-sided 97.5% lower confidence bounds as an alternative, this is also acceptable.

V. LABELING FOR WAIVED DEVICES

In order to assure that the device is "simple" and has "an insignificant risk of an erroneous result," your labeling should contain instructions for running and interpreting the test accurately, written at a level appropriate for the intended operator. You should include your proposed labeling, including Quick Reference Instructions, package insert, and outer labels, in your waiver application. Labeling for in vitro diagnostic devices must meet all applicable labeling requirements as stated in 21 CFR 809.10(b). In addition, the package insert for waived test systems should include additional information appropriate for untrained operators, such as the items listed below. Also, see Appendix A for more detailed recommendations.

- Identify the test system as waived and notify operators that facilities performing testing must have a CLIA certificate of waiver. 42 USC 263a(c)(2). Also, note that all applicable state and local laws must be met.
- Include a statement that laboratories eligible for a certificate of waiver must follow the test system instructions , including use with only the waived specimen type(s), instructions for limitations/intended use, and performance of QC testing as a failure-alert mechanism. 42 CFR 493.15(e).
- Integrate instructions for QC with procedural instructions for performing the test, in both the package insert and the Quick Reference Instructions.
- Include results of clinical studies that supported test waiver (in consultation with FDA). The performance information in your labeling can be finalized after study results are reviewed and the test is determined to be waived by FDA.

A. Quick Reference Instructions (QRI)

You should include Quick Reference Instructions (QRI) in your application. Preferably, the QRI should be laminated and attached to the test system. These instructions should be clear, easy to understand, in a readable font of 12 or greater, and include pictures, when possible. You should write instructions at no higher than a 7th grade reading level. For recommendations on what to include in the quick reference instructions, see Appendix A. You should include all the items in that appendix that are applicable to your test system, as well as any other appropriate information specific to your test system.

B. Quality Control (QC) Labeling Recommendations

Instructions should clearly explain why QC is needed and emphasize the value of external control testing at regular intervals for ensuring operator competency, and reagent and instrument (when appropriate) integrity. Instructions on how to perform control procedures using external controls are always recommended, and are critical if you are using them as a failure alert to help assure "insignificant risk of an erroneous result". Instructions relating to procedures used for QC should be integrated within the instructions for performing the test and should include the following:

- Step by step information on how to test control material, including testing frequency and concentration of materials.
- How to interpret results of control procedures.
- How to determine if results are invalid, (for example, for tests with an internal procedural control line, the test is not valid if the line is not present).
- Actions to take when results of control procedures are out of range or invalid.
- The limitations on all control mechanisms, including procedural controls, that you identified during the hazard analysis.

Your explanations of QC systems should include a description of what is being measured by all elements of both internal and external quality controls for a particular test system. To aid in addressing QC problems, you should provide a toll-free telephone number for technical assistance. We recommend that QC instructions take into account information obtained during the clinical studies (Section IV), as well as results of flex studies and validation studies under conditions of stress (Section III).

You should include discussions of benefits and limitations of the various device controls. For example, for a unitized test the following may be appropriate, and could be indicated in bold in the labeling for emphasis:

> "Test (*xyz*) contains built-in control features that monitor device functions (e.g., the presence of the control line shows that sufficient capillary flow has occurred). Obtaining the correct reading on the built-in control indicates that sufficient sample volume was added, but does not necessarily mean that your patient result is correct, because the built-in control does not monitor the entire assay. Good laboratory practice recommends the use of external positive and negative controls to assure the test reagents are working properly and that the operator has performed the test correctly. If the controls do not perform as expected, review the instructions for use to see if the test was performed correctly; repeat the test or contact technical assistance before testing patient specimens."

Examples of possible minimum frequency recommendations for running controls are as follows. You should base *your specific* recommendations on data from your studies.

- Each new lot.
- Each new shipment of materials even if it is the same lot previously received.
- Each new operator (we define this as an individual who has not run the test in the past 2 weeks).

- Monthly, as a check on continued storage conditions.
- Whenever problems (storage, operator, instrument, or other) are identified.
- If otherwise required by your laboratory's standard QC procedures.

C. Educational Information

As part of an overall plan to ensure that likelihood of erroneous results by the user is negligible, manufacturers should ensure that laboratories can read and understand the labeling, including test performance and limitations (as reflected by the analytical[5] and clinical studies) . We encourage you to consider innovative mechanisms to provide technical assistance to laboratories, and to ensure they understand the labeling information. We also recommend that manufacturers assist laboratories performing waived tests to become better educated on proper laboratory techniques. For example, we recommend that you participate in the development and promotion of good laboratory practice guidelines by developing training and education programs for the end operator. We also encourage you to incorporate proficiency testing, when feasible, and to promote laboratory participation in proficiency programs. In addition, we recommend that you include good laboratory practice information in the package insert, in accessory educational or technical material, and through the development of formal educational training programs. You should provide information on the following topics to operators:

- Importance of retaining a current version of the package insert.
- Importance of following the test instructions in the sequence given in the instructions.
- Need for proper operator training and retraining in order to maintain competency.
- Need for users to follow all instructions related to storage, preparation, and expiration dating, in order to maintain adequate test performance.
- Importance of maintaining procedures and results, as needed for proper performance of the test and patient management.
- General purpose of quality control, and value of using quality control within a broader system of quality assurance.

VI. SAFEGUARDS FOR WAIVED TESTS

1. In order to help ensure that the waived device will have "an insignificant risk of an erroneous result" after the product is marketed to waived settings, FDA is recommending that manufacturers of waived tests put a brief description of the MedWatch medical products reporting program along with the MedWatch phone number (1-800-FDA-1088), and fax numbers (1-800-FDA-0178), and website (www.fda.gov/medwatch) in the package insert. You may also describe how the MedWatch program works, which failures should be reported to both the company and FDA, and when failures should be reported to ensure proper tracking and reporting of waived testing issues.

[5] Analytical studies refers to those reviewed in the premarket application

2. Manufacturers of waived devices must maintain and implement medical device reporting procedures as required by 21 CFR 803.17 and must establish and maintain medical device report (MDR) event files as required by 21 CFR 803.18. See also 21 CFR 803.20, 21 CFR Part 803, Subpart E.

3. Manufacturers of waived devices must submit MDRs of individual adverse events as required by 21 CFR 803.10(c). We recommend manufacturers also notify CMS when device failures are reported.

VII. REFERENCES

1. ISO 17511: 2003 (E) In vitro diagnostic medical devices – measurement of quantities in biological samples – metrological traceability of values assigned to calibrators and control materials.
2. CLSI (former NCCLS). User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline. CLSI document EP12-A (ISBN 1-56238-468-6). CLSI, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2002.
3. CLSI (former NCCLS). Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline. CLSI document EP21-A (ISBN 1-56238-502-X). CLSI, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003.
4. CLSI (former NCCLS). Method Comparison and Bias Estimation Using Patient Samples; Approved Guideline – Second Edition. CLSI document EP9-A2 (ISBN 1-56238-000-0). CLSI, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2000.
5. CLSI (former NCCLS). How to Define and Determine Reference Intervals in the Clinical laboratory; Approved Guideline – Second edition. CLSI document C28-A2 (ISBN 1-56238-406-6). CLSI, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2000.
6. Fisher LD, Van Belle G. *Biostatistics. A Methodology for the Health Sciences* 1993, John Wiley & Sons, Inc.
7. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method comparison analysis. *Clinical Chemistry* 1979; 25:432-438.
8. Linnet K. Evaluation of the linear relationship between the measurements of two methods with proportional errors. *Statistics in Medicine* 1990; 9:1463-1473.
9. Martin RL. General Deming Regression for Estimating Systematic Bias and Its Confidence Intervals in Method-Comparison Studies. *Clinical Chemistry* 2000; 46(1):100-104.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307-310.
11. Parkes, JL, Slatin, SL, Pardo S, Ginsberg, BH. A new Consensus Error Grid to Evaluate the Clinical Significance of Inaccuracies in the Measurement of Blood Glucose. *Diabetes Care* 2000;23:1143-1148

12. Clarke WL, Cox, D, GonderFrederick LA, Carter W, and Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 1987;10(5):622-628.
13. Clopper CJ, Pearson E. The use of confidence or fiducial limits illustrated in the case of Binomial. *Biometrika* 1934;26:404-413.
14. Altman D.A., Machin D., Bryant T.N., Gardner M.J. eds. *Statistics with Confidence.* 2nd ed. British Medical Journal; 2000.
15. Agresti, A. and Coull, B.A. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* 1988;52:119-126.
16. Agresti, A. *Categorical Data Analysis*. 1990, John Wiley & Sons, Inc.

Appendix A

SPECIFIC LABELING RECOMMENDATIONS FOR WAIVED DEVICES
QUICK REFERENCE INSTRUCTIONS (QRI) WITH PICTURES AND DIAGRAMS should generally contain the following and any other sections appropriate for your specific device. The QRI should be written at no higher than a 7th grade reading level:
The name of the test and statement that labs with a waiver certificate may use it
A statement that users should read the complete test procedure, including recommended QC procedures, before performing the test and that they should refer to the package insert for more complete information. If appropriate, a statement that users should perform control procedures before performing the test.
A statement that laboratories with a certificate of waiver must follow the manufacturer's instructions for performing the test. 42 CFR 493.15(e)(1).
Step-by-step test instructions. Include, as appropriate: physical environmental specifications/conditions for test performance; specifications for specimen collection, handling, storage and preservation; preparation of reagents and control materials; storage of reagents and control materials; and calibration procedures. Utilize diagrams and flowcharts to illustrate how to run the test, when helpful.
Step-by-step instructions for all control procedures including frequencies, action to be taken if control results are out of range, or invalid, or if other failure alert or fail safe mechanisms are activated.
Interpretation of results, including diagrams on how to read and assess validity of test results and control results.
A warning addressing color blindness when waived tests use color-coded reagents and/or endpoints.
Safety considerations for test operation that particularly apply to untrained users.
Critical maintenance, such as cleaning (including safety considerations).
PACKAGE INSERT - Considerations for waived tests (in addition to any other requirements specified in 21 CFR 809.10 and any other considerations specific for your device type)
Identification of the test as CLIA waived, a statement that a certificate of CLIA waiver is required to perform the test in a waived setting, and information on how users can obtain a certificate.
A statement that laboratories with a certificate of waiver must follow the manufacturer's instructions for performing the test. 42 CFR 493.15(e)(1).
Test operation safety considerations that particularly apply to untrained users.
The physical environmental specifications/conditions for performing the test.
A warning addressing color blindness when waived tests use color-coded reagents and/or endpoints.
Step-by-step operating instructions for performing the test, integrated with instructions for all control procedures.
Action to be taken when no test result is obtained or when the results is out of the reportable range.

Study results demonstrating how the test compares to a known method, traceable to a reference method, if applicable.
A brief description and summary of the results from the waiver studies.
When appropriate, warnings about clinical errors that can occur even when the test result is analytically correct.
Instructions indicating when and how additional testing should be done (e.g., in cases where results should be confirmed by a reference procedure performed by an appropriately certified laboratory).
Any other limitations, restrictions, and special considerations for your test system.
Appropriate QC recommendations or requirements (see below, "Quality Control Labeling Recommendations").
Information on reporting test system problems to the manufacturer and/or FDA. You should include statements encouraging users to contact you and/or FDA so that you can track and account for device problems. (www.fda.gov/medwatch)
A statement that users should notify CMS of device problems. http://www.cms.hhs.gov/clia/ro-map.asp
Manufacturer contact information (phone number and whom to contact)
Quality Control Labeling Recommendations
Step by step information on how to test control materials.
Frequency for testing control materials.
How to interpret control results and procedural controls.
Actions to take when control results are out of range, or invalid.
Limitations of the device's controls, identified during the hazard analysis.

Appendix B DEFINITION OF TERMS AS USED IN THIS DOCUMENT

"Accurate"	In this guidance document we use the term "accurate" tests (in quotes) to refer to those tests that are comparable to a traceable method, in which the results of measurements can be related to stated references. We write the term "accurate" in quotes to denote that it is used in the context of CLIA and this guidance. The definition of accuracy may be different in other contexts, including those provided in other scientific standards, such as reference [1].
Allowable total error (ATE)	In this document, the allowable total error is the limit for the differences between the WM and the CM or the WM and the RM that can be tolerated without invalidating the medical usefulness of the test. These differences can be expressed as percents of CM values for high values, and as a concentration difference for low values. Typically at least 95% of the observed differences should fall within the established ATE zone.
Control material	Material used to verify performance characteristics of a medical device.[6]
Control procedures	Operational techniques and activities at the point of use to monitor the performance of the device and fulfill the laboratory's requirements for quality[6]. Any single control procedure might monitor all or part of the measurement procedure, from the collection of sample to reporting the result of the measurement.
External control material	Control material that is not built-in to the device. Typically this is in a similar matrix as the intended use specimen and is processed using the same procedures as patient specimens.[6]
Fail-safe mechanisms	Mechanisms to ensure that a test system does not provide a result when test conditions are inappropriate, or when the result is based on faulty test functioning.
Failure alert mechanisms	Mechanisms that notify the operator of any test system malfunction or problem. Failure alert mechanisms ideally include built-in controls or checks. Procedures that use external control material can also be considered failure alert mechanisms.
Flex studies	Studies performed using the device under conditions of operational stress. These studies are performed as part of the hazard analysis, to identify sources of error.

[6] Also see Harmonized Terminology Database at the CLSI website: www.clsi.org for further details and more general use of the term.

Hazard Analysis	A process used to identify device hazards, and the sources of error that may cause the hazard, as well as to implement and test mechanisms to mitigate the sources of error and reduce their risks to an acceptable level.
Hazards (for IVD's)	Potential source of harm (to a patient or test operator). For IVD's, hazards for patients are generally incorrect patient results or operator injuries
Intended operator (user)	In this guidance, intended operator (user) refers to a test operator with limited or no training or hands-on experience in conducting laboratory testing (e.g., medical assistant, nurse, doctor, or an individual with no medical training).
Internal control	A control, or system check, built into the device system, i.e., the user does not need to use additional reagents to perform that particular control process.
Laboratory Professional	A person who meets the qualifications to perform moderate or high complexity testing, such as a medical technologist (MT) or medical laboratory technician (MLT).
Limits of erroneous results (LER)	In this guidance the limits of erroneous results are limits for the differences between the WM (waiver method) and the CM (comparator method) or the WM and the RM (reference method). Results outside these limits pose a risk to patient safety.
Matrix	The totality of components within a. patient specimen, controls, or calibrator other than the analyte.
Matrix effects	The influence of a property of the sample, other than the analyte, on the devices performance characteristics. [6] Examples: (a) The test values for a particular analyte in whole blood may differ from those for a fingerstick. (b) Control material in a matrix different from that of the specimen should be tested to ensure that test performance is the same as that for the specimen.
Procedural control	Controls or indicators to monitor whether specific aspects of the procedure were performed correctly. Often, procedural controls are in the form of control lines on a single use cassette or dip devices, and indicate whether sufficient sample was applied. Procedural controls generally do *not* serve as a control for the entire test system.
Quality control (QC)	The entire set of procedures and system checks designed to monitor the test method and the results to assure acceptable test system performance. [6]
Quick Reference Instructions	A short (usually one to two page) version of the test instructions, preferably laminated, that can be posted. It contains instructions needed on a frequent basis and directs operators to the package insert for topics such as performance characteristics, long term maintenance instructions,

	troubleshooting, and other more detailed instructions.
Reference material	A preparation of the analyte whose concentration and purity are sufficiently well-established and well-recognized so that the material is suitable for calibration or value assignment. [6]
Reference method	A method which has been thoroughly investigated and has been shown to yield values having trueness and precision of measurement such that the method can be used to assess the trueness of other methods for the same analyte or for assigning values to reference materials. [6]
Source of error	A component of the device, measurement method, or operator practice that can lead to device failure, thereby creating a risk for patients, operators, or other individuals. [6]
Total analytical error	The combination of errors from all sources, systematic and random. It is often expressed in terms of an interval that contains a specified proportion (e.g., 95%) of the distribution of observed differences between the WM and CM values (for the CM of types A, B and C) as well as the difference between WM and RM values (for the CM of type C). Sometimes relative differences (e.g., ((WM-CM)/(CM)) are used instead of differences.
Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. The measure of trueness is usually expressed in terms of bias. It is also often referred to as systematic bias. [6]