Challenges in Applying CLIA Waiver Guidance for Quantitative Tests

Ken Kupfer, Ph.D.
Alere San Diego, Inc.
Principal Scientist
Research and Development

FDA / Industry OIVD Roundtable Meeting, June 10, 2011

Introduction

Major challenges in applying the CLIA Waiver Guidance [1] for quantitative tests:

(1) The construction of an error grid:

The error grid should reflect the clinical requirements based on the medical utility of the test. Different tests and indications have different medical decision thresholds (often multiple thresholds per test) which lead to different tolerances for error.

(2) The comparison to a Gold Standard comparative method:

The CLIA Waiver Guidance requires identification of an appropriate comparative method. What happens when a Gold Standard is not available?

[1] Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices Document issued on: January 30, 2008 (OMB control number: 0910-0598).

Outline

Examples of Constructing Allowable Total Error (ATE) zones:

- INR testing of patients on oral anticoagulation therapy (OAT).
- Total Cholesterol testing of patients at risk for coronary heart disease (CHD).
- hsCRP testing of patients at risk for CHD.
- Creatinine testing of chronic kidney disease (CKD) patients.
- BNP testing of heart failure (HF) patients.

The difficulty in choosing a Gold Standard comparative method:

- Significant variation of “Gold Standards” for INR testing.
- BNP testing, where there is no gold standard.

Blood Glucose Testing of Diabetic Patients: Clarke and Parkes error grids [Diabetes Care 23:1143–1148, 2000] are well established and accepted (not discussed further in this presentation).

Error Grid Definition

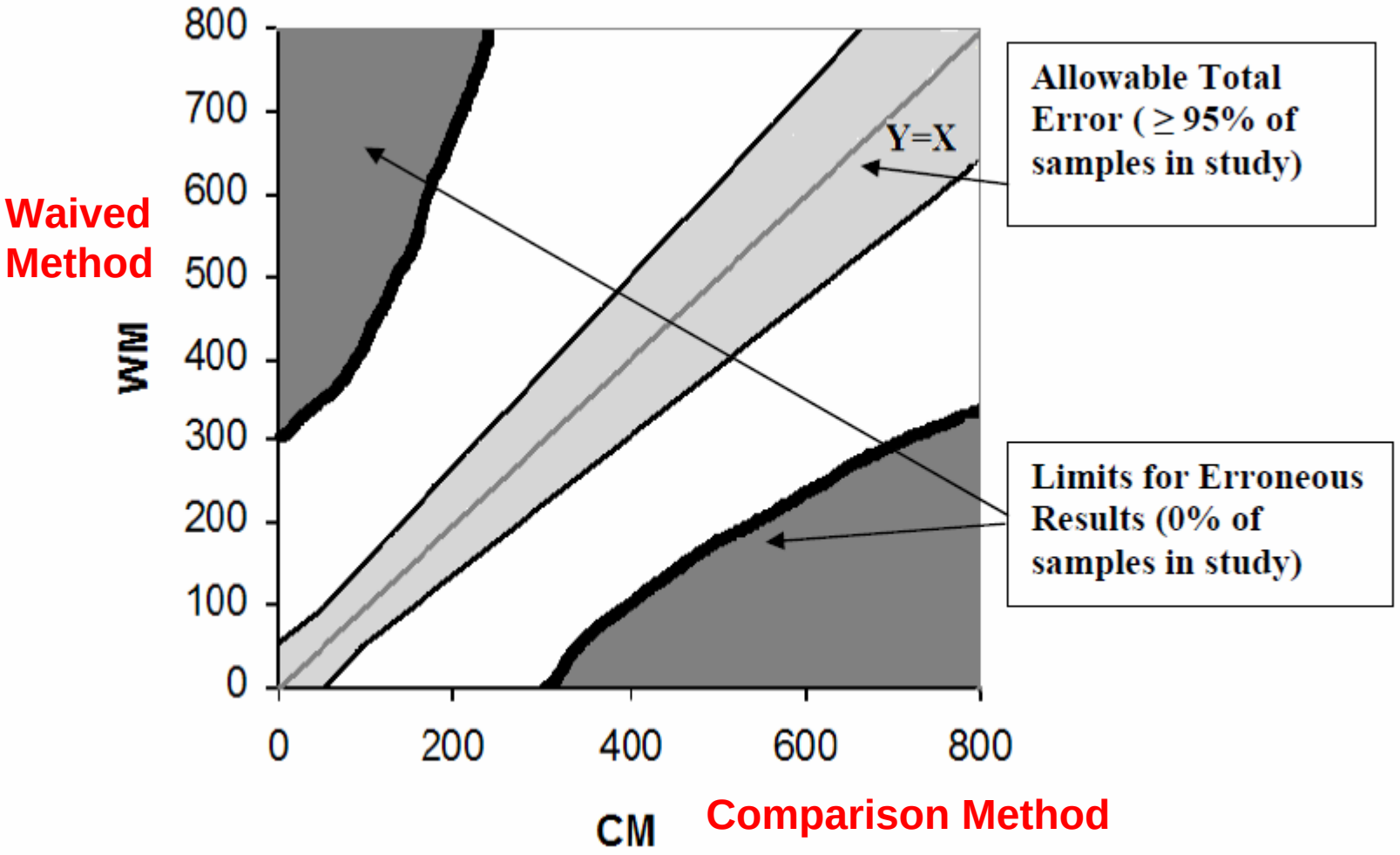


Figure 1. Example of ATE and LER zones

Figure from CLIA Waiver Guidance [1].

INR Testing

The ISO 17593 Guidance [2] sets a widely accepted standard for INR self-testing of patient's on oral anticoagulation therapy:

90% of the test results should be within the following parameters:

- Below 2.0 INR, results should be within ± 0.5 INR.
- From 2.0 through 4.5 INR, results should be within ± 30%.
- At levels higher than 4.5 INR, there are no requirements for accuracy.

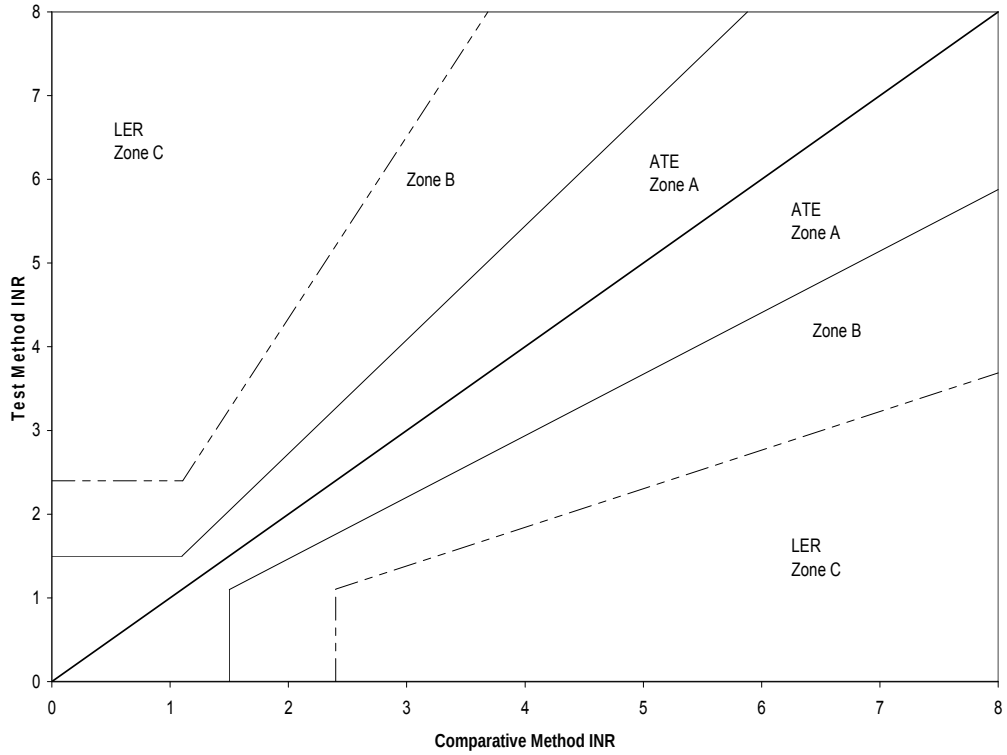
The CLSI EP27 Guideline [3] gives an example of an error grid based on the above tolerances (from the ISO 17593 Guidance).

[2] Clinical laboratory testing and in vitro medical devices — Requirements for in vitro monitoring systems for self-testing of oral anticoagulant therapy, ISO 17593:2007(E).

[3] How to Construct and Interpret an Error Grid for Diagnostic Assays; Proposed Guideline, Clinical and Laboratory Standards Institute, EP27-P.

INR Error Grid

The figure below is from the CLSI EP27 [3].
The ATE zone for INR > 1.5: 95% of the data must be within ± 36%*.



*The ISO specification is that 90% of the data lie within +/- 30%. To convert this to a limit encompassing 95% of the data, multiply 30% by z(97.5%)/z(95%), where z(p) is the quantile of the standard normal cumulative distribution for a given probability. Since z(97.5%) = 1.96 and z(95%) = 1.65, the result is that 95% of the data must lie within +/- 36%.

Why is ± 36% a Rational ATE for INR Testing?

An individual on OAT has a target therapeutic INR range that generally falls within the INR interval of 2.0 to 4.5:

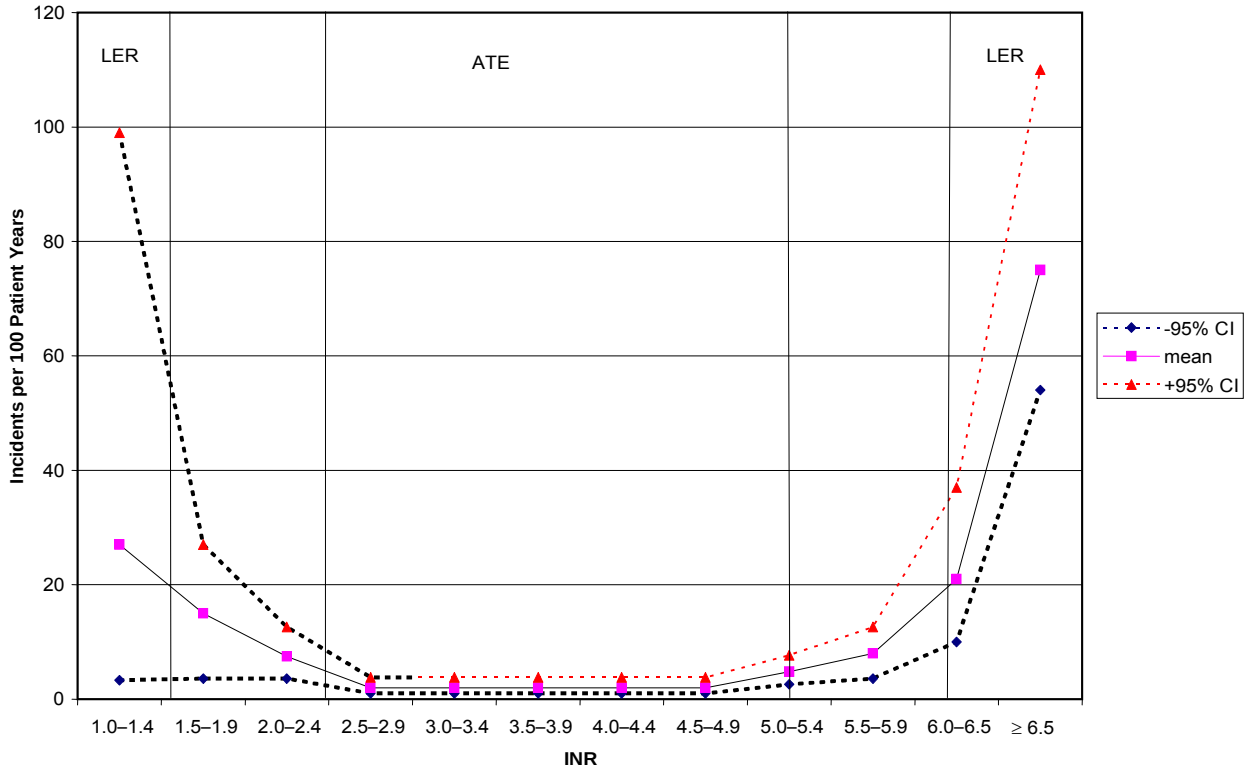


Figure from CLSI EP27 [3], originally from Cannegieter SC, Rosendaal FR, Wintzen AR, van der Meer FJ, Vandenbroucke JP, Briët E. Optimal oral anticoagulant therapy in patients with mechanical heart valves. N Engl J Med. 1995;333(1):11-17.

If patients are below 2.0, or above 4.5, then physicians may adjust dosage. INR within the 2.0 to 4.5 interval would typically not result in a change of dose. An INR > 4.5 relative to a true INR < 2.0 (or vice versa) may result in a change of dose that is opposite to the patient's actual need.

Is there a Formula for ATE?

A simple formula for the case of a test with two decision thresholds flanking a “grey zone” (CLSI Guideline [3] explains the logic, but does not give the formula):

X = marker concentration
L = lower threshold for medical decision
H = upper threshold for medical decision

Patients with $L \leq X \leq H$ are treated similarly, i.e., not distinct with regard to medical care, or decision making. This is the “grey zone”.

Set ATE limits such that a concentration at **the mid-point of the grey zone has test results within the grey zone 95% of the time** assuming normally distributed errors:

$$\mathbf{ATE = \pm (H-L)/(H+L)}$$

Example for INR Testing of patients on OAT:

- Low threshold is 2.0
- High threshold is 4.5
- ATE (from formula) is equivalent to ±1.25 at an INR of 3.25, i.e., ± 38%.
- Agrees with the ATE (36%) in the CLSI Guideline [3], consistent with ISO [2].

A rational ATE depends on the test and the indication: it may be narrow, or wide:

In the previous example (INR testing), an ATE of ± 36% is appropriate. In the following example (Total Cholesterol Testing), an ATE of ± 8.9% is appropriate.

Total Cholesterol Testing of patients at risk for coronary heart disease (CHD):

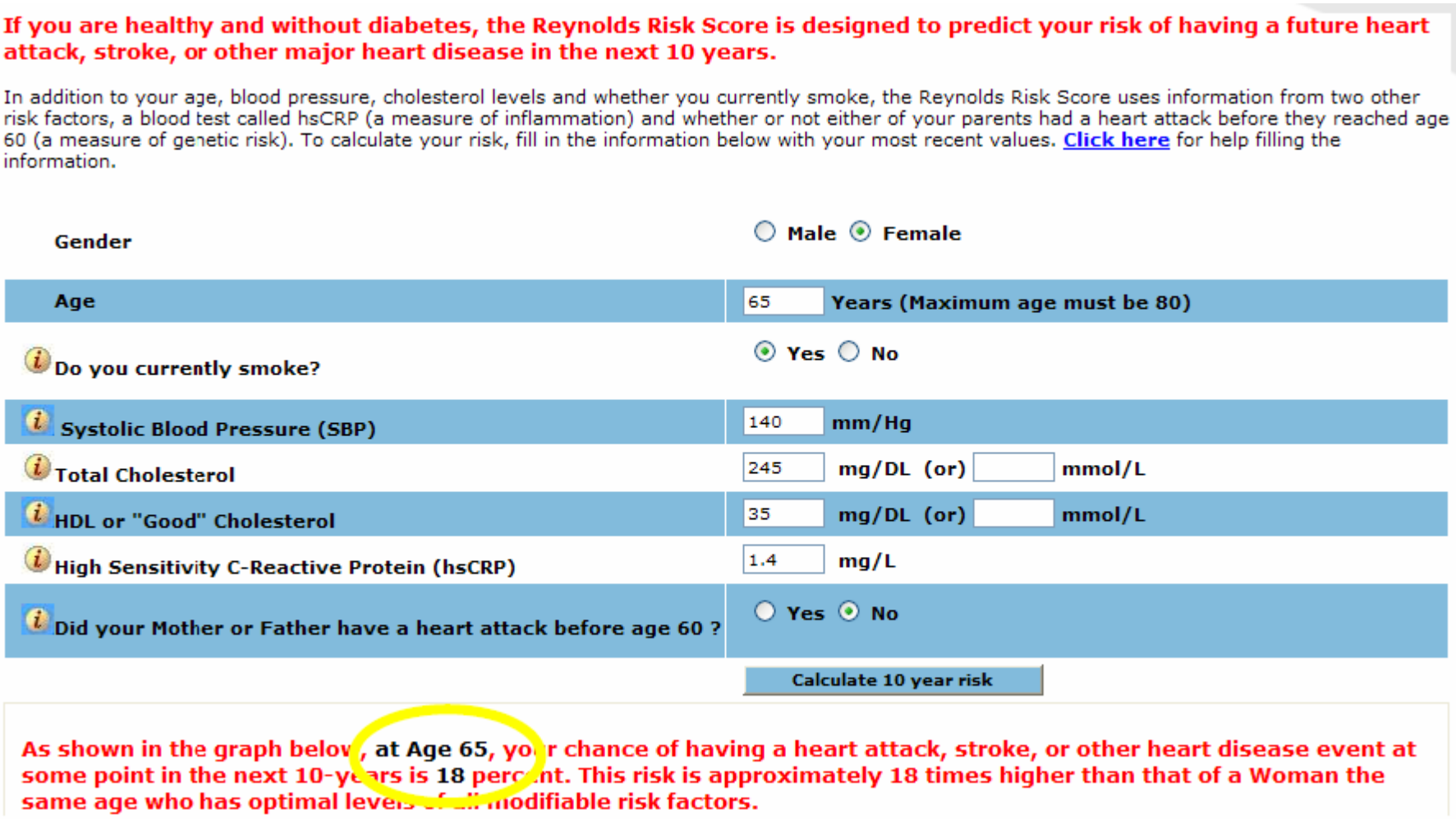
- Low risk is indicated by values below 200 mg/dL
- High risk is indicated by values above 240 mg/dL
- ATE (from formula) is ± 20 mg/dL at 220 mg/dL, i.e., ± 9%.
- The NCEP Guideline [4] sets ATE at ± 8.9%.

So different tests and indications have different medical decision thresholds (multiple thresholds per test) which lead to different tolerances for error.

[4] National Reference System for Cholesterol, Cholesterol Reference Method Laboratory Network, Total Cholesterol Certification Protocol for Manufacturers, October 2004, National Cholesterol Education Program (NCEP).

hsCRP testing of patients at risk for CHD:

Below is a snapshot of the Reynolds Risk Score Calculator* with an example parameter setting incorporating hsCRP together with other risk factors (graphical output omitted):



If you are healthy and without diabetes, the Reynolds Risk Score is designed to predict your risk of having a future heart attack, stroke, or other major heart disease in the next 10 years.

In addition to your age, blood pressure, cholesterol levels and whether you currently smoke, the Reynolds Risk Score uses information from two other risk factors, a blood test called hsCRP (a measure of inflammation) and whether or not either of your parents had a heart attack before they reached age 60 (a measure of genetic risk). To calculate your risk, fill in the information below with your most recent values. Click here for help filling the information.

Gender	○ Male ◉ Female
Age	65 Years (Maximum age must be 80)
Do you currently smoke?	◉ Yes ○ No
Systolic Blood Pressure (SBP)	140 mm/Hg
Total Cholesterol	245 mg/DL (or) mmol/L
HDL or "Good" Cholesterol	35 mg/DL (or) mmol/L
High Sensitivity C-Reactive Protein (hsCRP)	1.4 mg/L
Did your Mother or Father have a heart attack before age 60 ?	○ Yes ◉ No

Calculate 10 year risk

As shown in the graph below, at Age 65, your chance of having a heart attack, stroke, or other heart disease event at some point in the next 10-years is 18 percent. This risk is approximately 18 times higher than that of a Woman the same age who has optimal levels of all modifiable risk factors.

*http://www.reynoldsriskscore.org/home.aspx

hsCRP testing of patients at risk for CHD:

hsCRP Testing of patients at risk for coronary heart disease (CHD):

- **Low risk is indicated by values below 1 mg/L**
- **High risk is indicated by values above 3 mg/L**
- ATE (from formula) is ± 1.0 mg/L at 2.0 mg/L, i.e., ± 50%.

There is no guidance on the ATE for this application. However, the above suggests an ATE of ± 50% using the **same logic** as for INR testing and also for Total Cholesterol testing (each of which satisfies the relevant guideline).

Why can't a first generation **Waived Method** take advantage of the full ± 50% implied by the medical decision points?

Creatinine Testing of Patients with CKD:

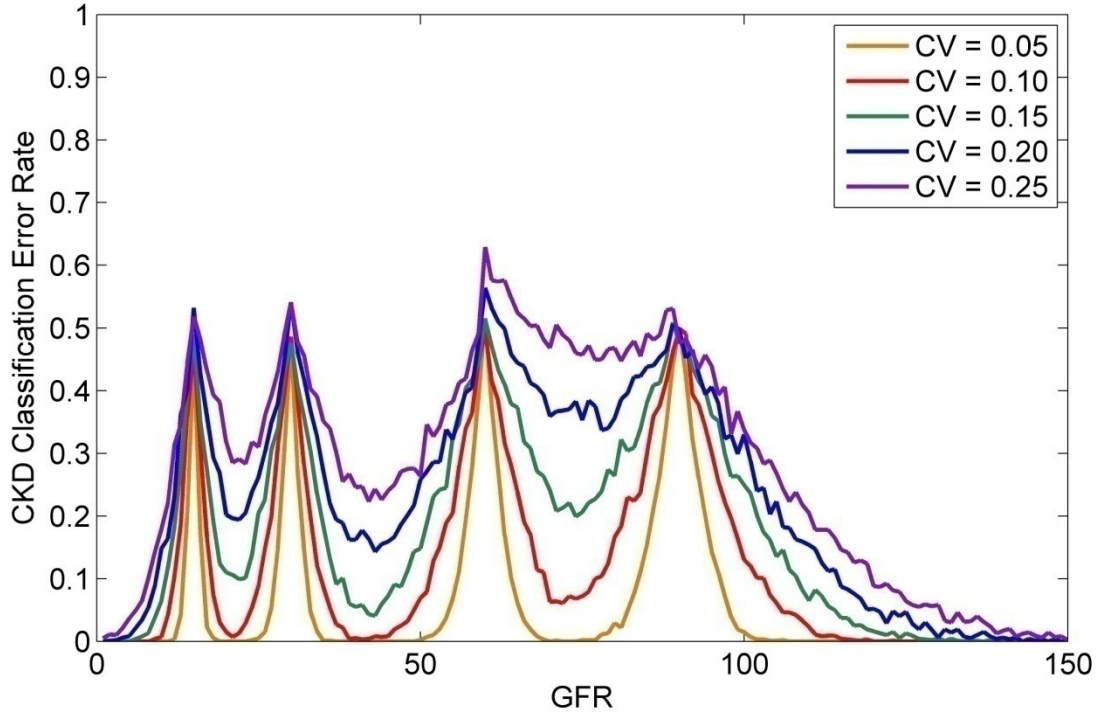
Risk and Progression of CKD [5] are tracked over time (months to years) by **estimating** Glomerular Filtration Rate (GFR), typically by **measuring** creatinine and adjusting for age, gender, and race (MDRD equation [6]).

CKD Stage	GFR mL/min/1.73 m^2	Description
1	GFR≥90	Normal kidney function but urine findings or structural abnormalities or genetic trait point to kidney disease
2	60≤GFR<90	Mildly reduced kidney function, and other findings (as for stage 1) point to kidney disease
3	30≤GFR<60	Moderately reduced kidney function
4	15≤GFR<30	Severely reduced kidney function
5	GFR<15, or on RRT	Very severe, or endstage kidney failure (sometimes call established renal failure)

[5] National Kidney Foundation. *KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification.* Am J Kidney Dis 39:S1-S000, 2002 (suppl 1)
[6] GFR = 175 · $sCr^{-1.154}$ · $age^{-0.203}$ · 1.212 (if patient is black) · 0.742 (if female),
Using Standardized Serum Creatinine Values in the Modification of Diet in Renal Disease Study Equation for Estimating Glomerular Filtration Rate, Levey et al, Annals of Internal Medicine, 2006.

Creatinine Testing of Patients with CKD:

Misclassification of CKD Stage based on simulations* of estimated GFR with measurement error relative to the assumed true value.



*Simulations of estimated GFR at a given "true" GFR (1000 simulations at each GFR from 1 to 150) assuming that GFR α 1/X, where X is a Creatinine measurement with normally distributed errors (SD = X·CV).

Based on the spacing of the CKD stages, we argue that the required ATE should be ± 20%. For example, the "grey zone" for Stage 2 is a GFR interval of 60 to 90, so the required resolution for measurement is GFR = 75 ±15.

BNP Testing of HF Patients

Aid in the Diagnosis of Heart Failure (HF):

- 100 pg/mL BNP is a well established cutoff† for aid in the diagnosis of acute heart failure in patients with ambiguous signs and symptoms, and as an aid to exclude the diagnosis of heart failure in the non-acute setting
- A gray zone of 100-500 pg/mL is commonly used in clinical practice where <100 pg/mL is considered for rule-out diagnosis of heart failure and >500 pg/mL is considered for rule-in‡

HF severity assessment and patient management:

- In addition to signs and symptoms, an approximate doubling of BNP concentration is indicative of a worsening condition in patients with heart failure due to the high intra-individual biological variation of BNP
- NACB and IFCC Guidelines recommend that caution be exercised in interpreting <50% changes in BNP concentration as being related to medical therapy because a high biological variation of BNP exists* (intra-individual biological variation in BNP is 30-40% based on meta-analysis of published reports).
- In patients hospitalized with Acute Decompensated Heat Failure, elevated discharge BNP is prognostic of poor outcomes (mortality and re-hospitalization).

†Maisel et al. N Engl J Med 2002: 347(3): 161-167.
‡ Coste et al. Clin Chem 2006: 52(12): 229-2235.
* Apple et al. Circulation 2007: 116: e95-e98.

BNP Testing of HF Patients

For Aid in the Diagnosis of HF, we would argue for ATE limits based on the established medical decision thresholds:

- Low threshold is 100 pg/mL
- High threshold is 500 pg/mL
- Total Error should not exceed ± 200 pg/mL at 300 pg/mL, i.e., ± 67%.

The medical decision thresholds for BNP testing for HF severity assessment and patient management are not well established. However, we may argue for ATE based on detecting a 2-fold change. For example, a BNP of 200 pg/ml implies the same treatment as a BNP is in the range of 100 pg/ml up to 400 pg/ml.

- Low threshold is 100 pg/mL
- High threshold is 400 pg/mL
- Total Error should not exceed ± 150 pg/mL at 250 pg/mL, i.e., ± 60%.

So we would argue for similar ATE (± 60%)* for either application.

*If errors were log-normally distributed, then asymmetric ATE limits of +100% and -50% may be appropriate.

Outline

Examples of Constructing Allowable Total Error (ATE) zones:

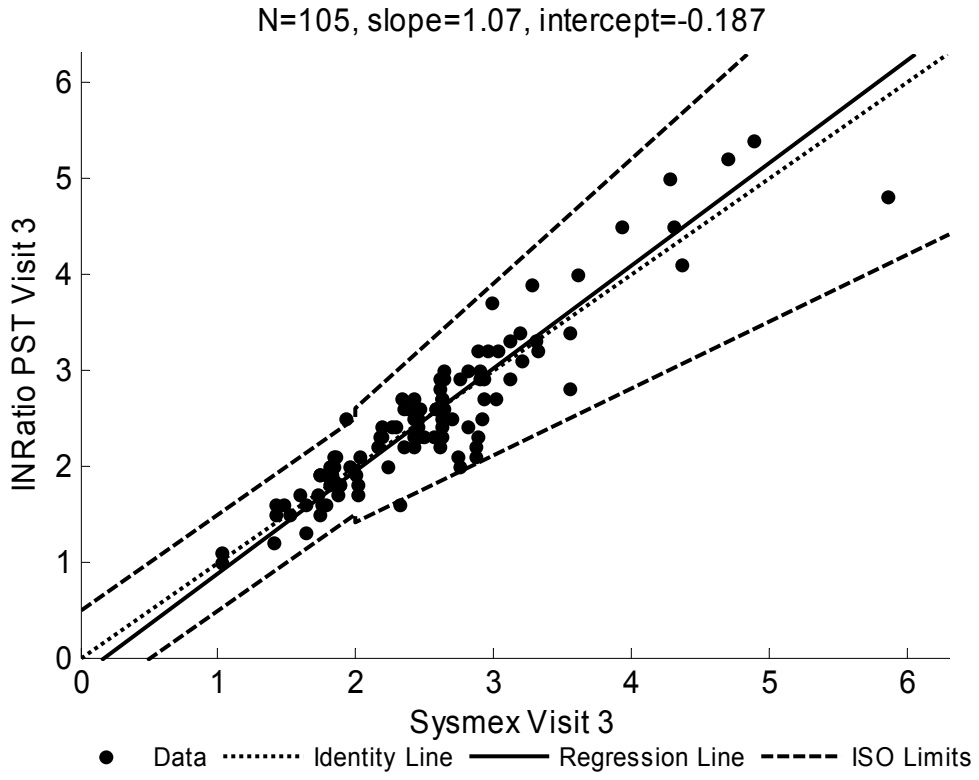
- INR testing of patients on oral anticoagulation therapy (OAT).
- Total Cholesterol testing of patients at risk for coronary heart disease (CHD).
- hsCRP testing of patients at risk for CHD.
- Creatinine testing of chronic kidney disease (CKD) patients.
- BNP testing of heart failure (HF) patients.

The difficulty in choosing a Gold Standard comparative method:

- Significant variation of “Gold Standards” for INR testing.
- BNP testing, where there is no gold standard.

INR Testing, Real Data Shows Significant Variation Between Methods

Data comparing Patient Self Testing (INRatio PST) on fingerstick whole blood versus a laboratory method (Sysmex) on plasma sample. Limit lines from ISO 17593 Guidance [2].



Between-Method CV (%)	Percent Outside ±36% ATE *
10	0.23%
11	0.46%
12	0.80%
13	1.28%
14	1.91%
15	2.70%
16	3.64%
17	4.72%
18	5.92%
19	7.22%
20	8.61%

Note, Laboratory Methods also show a Between-Method CV of 10-15%.

***Assumptions for "Percent Outside ATE" Table:**
Within-Method CV of the Comparative Method (plasma replicates) =2%
Within-Method CV of the Test Method (fingerstick duplicates) = 6%
Normally Distributed Error with SD equal to the square-root of the sum of the within-method and between-method variances.
Between Method CV includes sources of variance associated with sample matrix (blood-to-blood).

BNP Testing, When there is No Gold Standard

Given the clinical requirements (implied by well separated medical decision thresholds), a patient self-test for INR can still meet the ATE goal, despite the challenge of between-method variance.

For BNP there is no recognized gold standard comparative method or reference materials; moreover, because of sample matrix effects resulting from differential binding of antibodies used in various methods to the actual BNP antigen and to some of its degradation products and precursor molecules, comparison to another method does not tell what method gives true BNP.

Assuming that a fingerstick BNP system (strip and meter) has been cleared for professional use, then why not evaluate CLIA waiver based on a direct comparison of operators, waived vs. professional, using the same sample type at the same time on the same method?

Conclusions

(1) Working within the constraint of the CLIA Waiver Guidance, we should set the ATE based upon rational criteria, reflecting the clinical requirements based on the medical utility of the test.

(2) When there is no Gold Standard, a direct comparison of professional and waived operators (using the same method) may be appropriate.