

MDICx: Developing Clinical Evidence for Regulatory and Coverage Assessments in *In Vitro* Diagnostics

Tuesday, August 13, 2019



Align. Achieve. Accelerate.

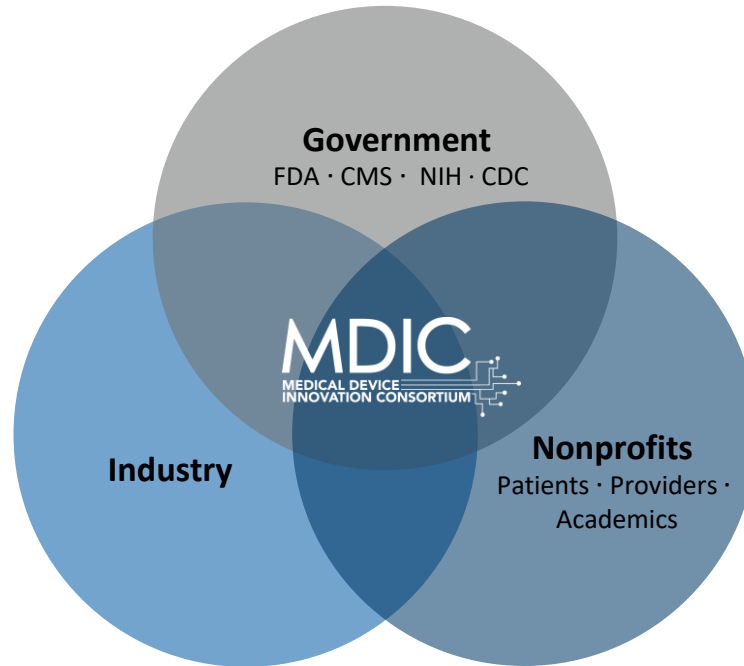
Panelists:

- *Jaime Houghton*, MS, MT(ASCP) | Sr Director, Regulatory Affairs, Sysmex Americas
- *Louis B. Jacques*, MD | Chief Clinical Officer, ADVI
- *Timothy Stenzel*, MD, PhD | Director OHT7: Office of In Vitro Diagnostics and Radiological Health | CDRH | FDA

MDIC Overview

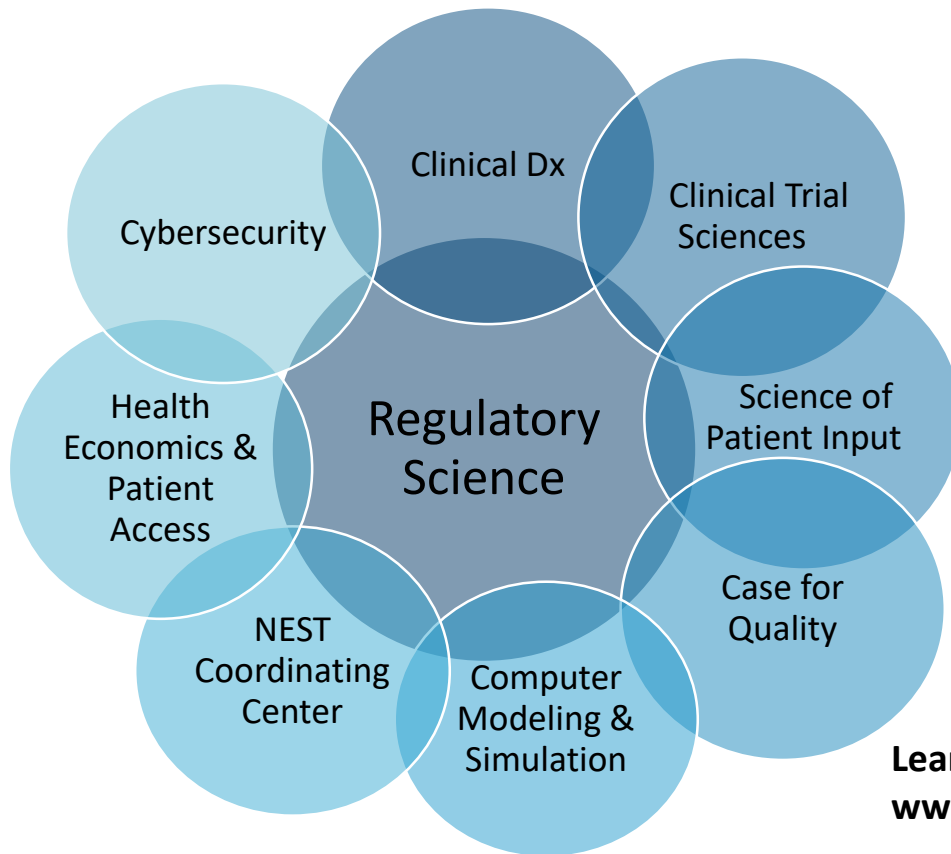
Carolyn Hiller, MDIC Clinical Diagnostics Program Director

MDIC is a 501(c)(3) non-profit organization and is the first-ever public-private partnership created with the sole objective of advancing regulatory science of medical devices for patient benefit



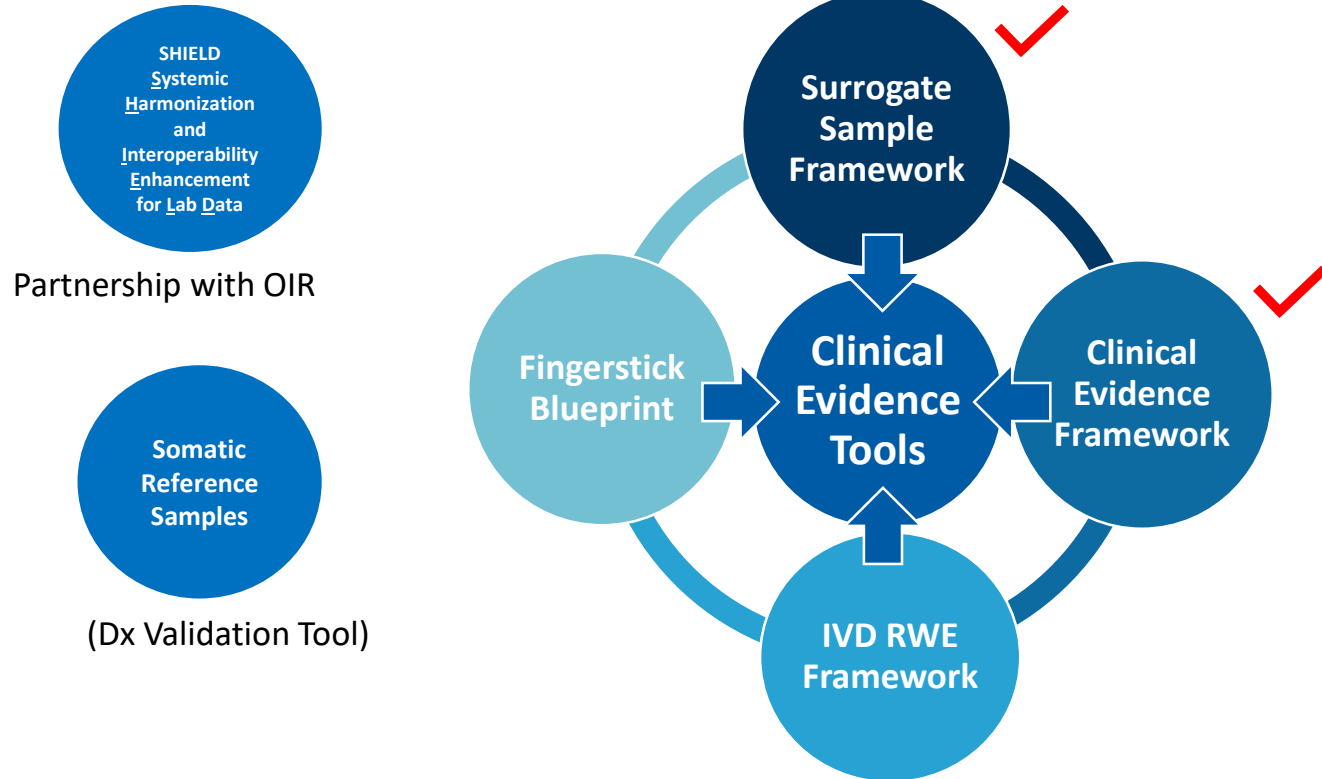
Resources · People · Intellectual Capital

MDIC PROGRAM INITIATIVES



Learn more at:
www.MDIC.org

Clinical Diagnostic Projects



Framework Sections

The background features a series of overlapping, semi-transparent geometric shapes in various shades of teal and light blue. Two thin white lines originate from the top left and extend towards the right, each ending in a small white dot.

Sections of the Framework:

- 1) Introduction
- 2) Analytical Validity
- 3) Clinical Validity
- 4) Clinical Utility
- 5) References

Analytical Validity

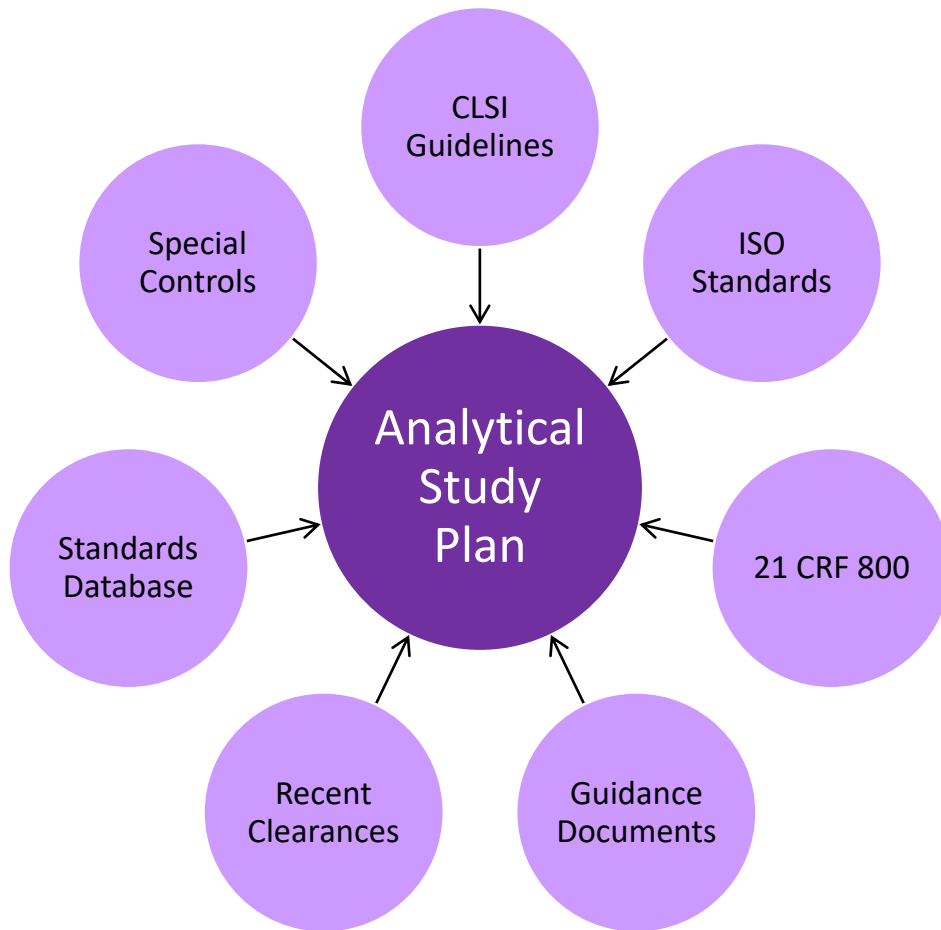
Jaime Houghton

What is Analytical Validity?

Ability of a test to **accurately** and **reliably** measure or detect the analyte(s) of interest in specimens that are representative of specimens that would be obtained in the intended use population

How is Analytical Validity Demonstrated?

Provide valid scientific evidence that the assay results are accurate, repeatable and reproducible, specific and sensitive



Analytical Studies

Accuracy

Precision

Linearity

Detection Capability

Analytical Specificity

Reference Interval

Measuring Interval

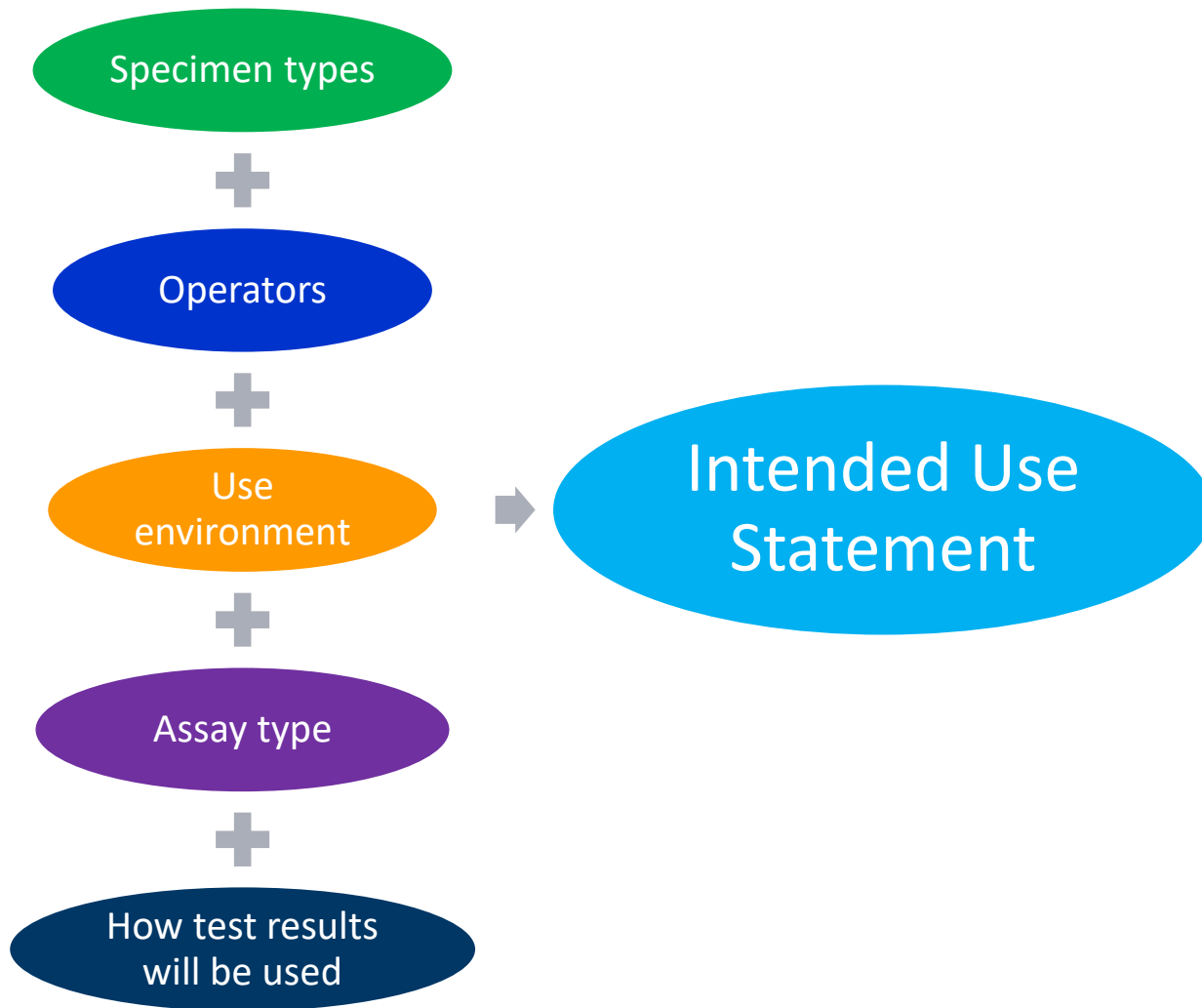
Carry-Over

Matrix & Specimen Type

Specimen Collection & Stability

Reagents, Cals & QC Stability

Usability & Flex Studies



Accuracy



- Closeness of agreement of the measured quantity value from the candidate device compared to the true value (comparator method)
 - Samples to cover entire analytical measuring range
 - FDA cleared device or standard or reference method as comparator

Precision



- Closeness of agreement between measured quantity value obtained by replicate measurements on the same or similar objects under specified conditions
 - Repeatability – same operating conditions over a short interval of time
 - Same operators, location and operating conditions
 - Reproducibility – reproducible operating conditions over time
 - Different operators, location and instruments

Linearity

- Ability, within a given range, to obtain test results that are directly proportional to the concentration of the analyte in the test sample
 - Seven to eleven samples of known concentration or known relative relationship to each other
 - Highest concentration above the upper limit and lowest concentration below the lower limit of linearity interval
 - Multiple replicates tested at each level

Detection Capability

- Lowest concentration of a measurand that can reliably be measured
 - Limit of Blank (LoB)- highest measurement result observed for a blank (negative) sample
 - Limit of Detection (LoD)- lowest level of analyte that can be distinguished from the LoB and reliably detected
 - Limit of Quantitation (LoQ) – lowest level of analyte that can reliably be detected and meet predefined bias and imprecision goals

Analytical Specificity

- Ability to assess, unequivocally, the analyte in the presence of other components that are expected to be present in the sample
 - Assess impact of potentially interfering substances on assay results
 - Consider both endogenous and exogenous sources of potential interference

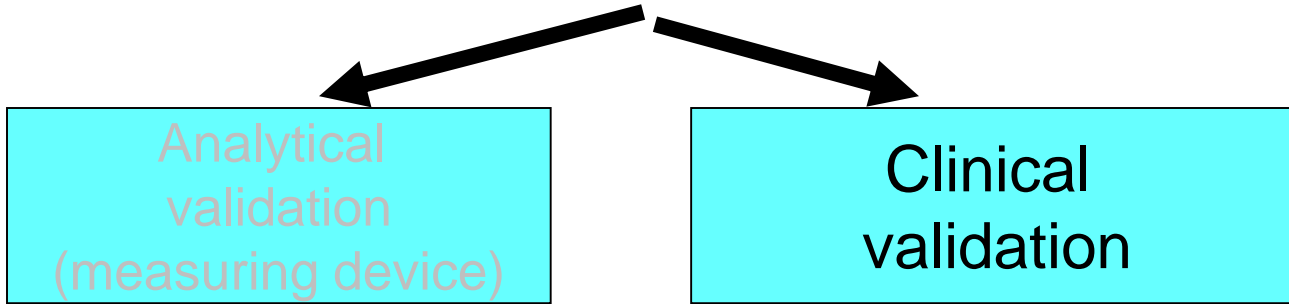
Reference Interval

- Interval between the lower reference limit and the upper reference limit of the subject population to be evaluated – “normal range”
 - Established by assessing approximately 120 apparently health subjects
 - Different age groups, gender groups and race/ethnicity groups may require establishment of different intervals

Clinical Validity

Timothy Stenzel

In Vitro Clinical Test



Analytical performance—does the test measure (detect) the analyte I think it does? Correctly? How reproducibly?

↓
Precision
Limit of Blank,
Limit of Detection,
Limit of Quantitation,
Linearity,
.....

Clinical validation—is a patient test result associated with the expected clinical presentation of this patient?

Clinical Validation – the process through which one shows that test results are clinically meaningful, i.e., finding whether the test is able to detect or predict the disorder or condition of interest in the target population

“Old analytes”:
the clinical validity/utility
is well established



Analytical validation
may be all that is
required.

Example:

- Calcium in serum

“Novel” analytes or
combinations of
analytes



Clinical validation study
may be needed.

Example:

- New biomarker for
detection of colon
cancer

“Old analytes” and
New/Changed intended
uses



New clinical validation
study may be needed

Example:

- CRP
Initial: infection, tissue
injury, inflammation
New: risk of cardiac
disease

Intended Use/Indication For Use

- What device measures or detects
- What device reports (qualitative, quantitative, ...)
- Specimen types (e.g., serum, plasma, urine,)
- Target (intended use) population
- Intended users (e.g., “untrained” users, layperson (OTC), ..)
- Target condition (e.g, colon cancer, ..)
- Medical testing contexts (e.g., diagnosis, screening, monitoring, ..)

Intended Use/Indication For Use

Example:

Analyte

*MammaPrint® is a qualitative in vitro diagnostic test service, performed in a single laboratory, using the **gene expression profile** of fresh frozen breast cancer tissue samples to assess a patients' risk for distant metastasis.*

*The test is performed for **breast cancer patients** who are less than 61 years old, with Stage I or Stage II disease, with tumor size ≤ 5.0 cm and who are lymph node negative. The MammaPrint® result is indicated for use **by physicians** as a **prognostic marker** only, along **with other clinicopathological factors**.*

**Intended Use
Population**

**Indication
For Use**

Clinical validation study design depends on:

- Intended Use/Indication for Use statement;
- Expected clinical validation characteristics (e.g., clinical sensitivity, clinical specificity, predictive values)
- Novelty of the technology
- Availability of an appropriate method to establish the true clinical status
-

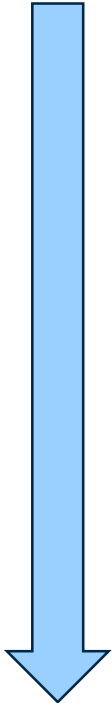
Examples of Medical Testing Contexts for IVDs

- ❑ **Diagnosis** (target condition is present or not during the time of testing);
- ❑ **Screening** (maybe in a general population (asymptomatic subjects at average risk) or a subpopulation (subjects at high risk))
- ❑ **Risk assessment** (assessment of predisposition to disease in future in asymptomatic patients)
- ❑ **Prognosis** (stratifying already diagnosed patients into poor or good prognosis)
- ❑ **Monitoring** (monitor the progress of disease or response to treatment)
- ❑ **Therapeutic response prediction**

Each context poses different risk and has its own requirements for clinical validity studies

Selection of Samples for Clinical Validation Study

Critical Component



Prospectively Obtained Samples

Archived Samples

Surrogate Samples

Prospectively Obtained Samples

- ❑ Pre-specified protocol for the clinical validation study
- ❑ Inclusion/exclusion criteria



Samples/subjects are
from the target population of the IVD test



Measures of the clinical performance (e.g., clinical sensitivity, specificity, ..) are **UNBIASED**.

Time of testing: maybe be collected and tested

- immediately (according to Instructions for Use)
- stored prior to being tested (according to the claimed sample stability)

Archived Samples

- ❑ Without pre-specified protocol or/and
- ❑ Protocol for different study



Additional Challenges

- Samples may be non-representative of the target population (e.g, left-over of large tumors, no samples near the cutoff because of re-testing, not enough sample volume, ...)
- Inclusion of samples positive by another IVD test
- Storage of samples (stability): especially challenging for new biomarkers; storage outside of the claimed sample stability

Archived Samples



Measures of the clinical performance (e.g., clinical sensitivity, specificity, ..) can be **BIASED**.

- ❑ If potentials for possible biases are adequately addressed, archived samples can provide unbiased estimates of the performance and therefore, can be used.
- ❑ Sometimes, archived samples are not representative (e.g., inclusion of samples positive by another IVD test) but they can be used in addition to prospectively collected samples (the best *available* source of *patient* samples).

Surrogate Samples

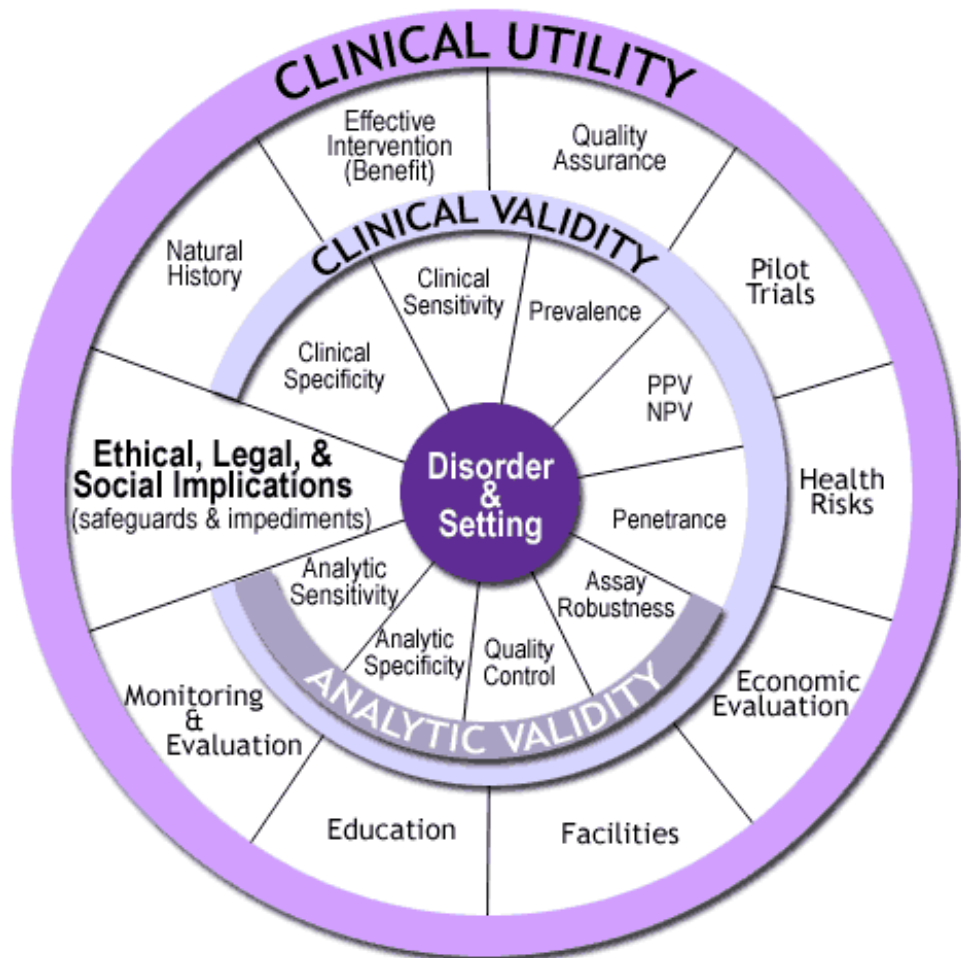
Sometimes, surrogate samples can be included in the clinical validation study (e.g., spiked samples close to the limit of detection)

Contact OIR *before* the use of archived and/or surrogate samples in the clinical validation study

Clinical Utility

Louis Jacques

ACCE



Clinical Utility is the Coverage Standard

Evidence of improved beneficiary health outcomes when the test is incorporated into management.

Cure is not required: test-based avoidance of futile burdens is a positive utility.

Can be direct or indirect chain of evidence.

- Direct = the clinical trial of the test assesses patient outcomes
- Indirect = the AV and CV of the test leverage existing therapy-based evidence of patient outcomes to establish a logical chain of evidence

Affirmed in Federal Court: *Kort v Burwell*.

Don't reinvent the wheel in a mature clinical paradigm, e.g. K⁺ management and cardiac rhythm disorders.



If you only need 3 numbers to open the lock (*i.e. manage the patient's condition*), why would it be reasonable or necessary to obtain 4 or 5 numbers?

Kort v Burwell, 2016

“As the court has concluded above, Defendants’ consideration of health outcomes and disease management was permissible under the Medicare Act itself.”

Why would CMS embrace clinical utility?


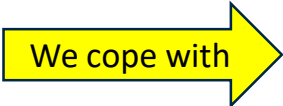
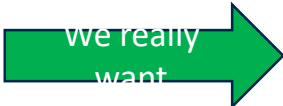
Well articulated paradigm:

- Well articulated in CDC, ACCE, EGAPP, SACGHS
- Lots of public input
- MEDCAC has endorsed it many times

Consistent with prior NCD language:

- Improved health outcomes
- Landmark paper by Fryback & Thornbury often cited in DMs
- Compatible with existing regulations (42CFR 410.32(a) about use by the treating physician.

Desirable Evidence

	Diagnostic Imaging Evidence Hierarchy Level	Genetic Testing Evidence Category	Example of Outcome Measures
	1. Technical Efficacy	1. Analytic validity	Interpretable scan resolution, accuracy and reliability of tests of CSF proteins to measure CSF protein levels, inter-reader and inter-laboratory reliability of test results
	2. Diagnostic Accuracy	2. Clinical validity	Sensitivity/specificity vs. gold standard test or vs. some other standard
	3. Diagnostic Impression		Change in presumptive diagnosis following introduction of new test results
	4. Diagnostic Action		Initiation or cessation of treatment; impact on use of additional diagnostic studies
	5. Patient Outcomes	3. Clinical utility	Cognitive/functional decline, time to institutionalization, side effects of treatment driven by test results, mortality
	6. Societal Outcomes		Cost-effectiveness of testing

The Rebuttable Presumption of Clinical Utility

“While a prospective trial with randomization and treatment based on the [TEST] score would solve these biases, and recognizing that long-term prospective data would require 10 years or more, Noridian and the CMS MoDX Contractor believe that clinical utility can be extrapolated from these robust retrospective clinical validity trials.”

Clinical Utility by Avoiding Futile Treatment

Decision Memo for Positron Emission Tomography (FDG) for Cervical Cancer (CAG-00181R2)

“As the requestors have written, the appropriate treatment strategy for cervical cancer is meaningfully changed if the patient is found to have distant metastases, in particular to the SUPRACLAVICULAR lymph nodes (Grigsby 2001, Tran 2003). The findings in several of the articles we have reviewed agree with the consensus of the MEDCAC: that, compared with other non-invasive methods, FDG PET is more sensitive in determining lymph node involvement in initial assessment of cervical cancer (Choi 2006; Selman 2008). In addition, publications support the beneficial effect on initial treatment planning of cervical cancer (Chao 2008, Hillner 2008), with **the majority of the effect being avoidance of futile surgery.**

“In summary, we conclude that the evidence is sufficient to determine that FDG PET for the initial staging of cervical cancer leads to improved physician management of this condition and **improves health outcomes** in Medicare beneficiaries who have been diagnosed with cervical cancer and is thus reasonable and necessary under Section 1862(a)(1)(A) of the Act. “

<https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=232&SearchType=Advanced&CoverageSelection=National&NCSelection=NCA&Keyword=supraclavicular&KeywordLookUp=Doc&KeywordSearchType=Exact&kq=true&bc=IAAAACAAQAAA&>

Fryback & Thornbury

Mol et al.

“We considered the evidence in the hierarchical framework of Fryback and Thornbury (1991) where Level 2 addresses diagnostic accuracy, sensitivity, and specificity of the test; Level 3 focuses on whether the information produces change in the physician's diagnostic thinking; Level 4 concerns the effect on the patient management plan and Level 5 measures the effect of the diagnostic information on patient outcomes. In evaluating diagnostic tests, Mol and colleagues (2003) reported: "Whether or not patients are better off from undergoing a diagnostic test will depend on how test information is used to guide subsequent decisions on starting, stopping, or modifying treatment. Consequently, the practical value of a diagnostic test can only be assessed by taking into account subsequent health outcomes." When a proven, well established association or pathway is available, intermediate health outcomes may also be considered. For example, if a particular diagnostic test result can be shown to change patient management and other evidence has demonstrated that those patient management changes improve health outcomes, then those separate sources of evidence may be sufficient to demonstrate positive health outcomes from the diagnostic test.”

This text is boilerplate in NCDs on diagnostic tests.



Questions?

Thank You!

<https://mdic.org/resource/framework-for-developing-clinical-evidence-for-regulatory-and-coverage-assessments-in-in-vitro-diagnostics-ivds/>

