WGME
Wessex Genomic Medicine Education

UNIVERSITY OF Southampton

Sensitivity and Specificity of Genomic Tests

Chris Mattocks
Wessex Regional Genetics Laboratory
27th February 2017

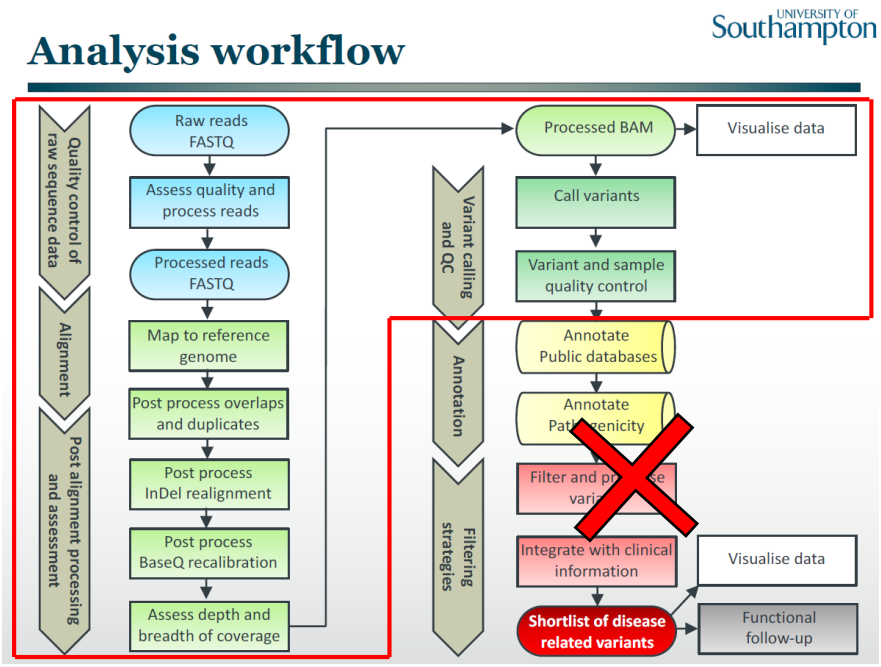
Outline

- Context
- Definitions
- Calculation
- Sampling error and reporting
- Example
- Key points
- Food for thought

Types of sensitivity

Analytical sensitivity (Biochemistry)

Technical sensitivity (Genetics)



Validation

- To ensure tests can robustly provide **ACCURATE** and appropriate results for patients
- Regulatory requirement of laboratory accreditation

The Standard

- **ISO 15189:2012 Medical laboratories -- Requirements for quality and competence**
  - Section 5.5 deals with pre-implementation validation and verification
  - Section 5.6 deals with on-going assurance of tests and defining measurement uncertainty

Definitions

- **Validation**: "Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" [ISO 15189: 3.26]

Is the test fit for purpose?

- **Verification**: "Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled" [ISO15189: 3.27]

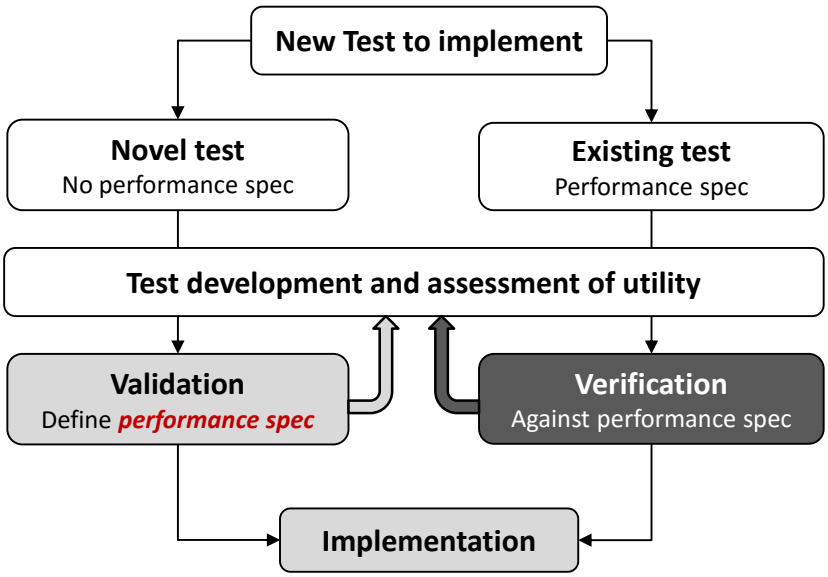
Is the *validated* test performing as it should?

When do we need to validate?

ISO 15189 – Examination procedures

- Tests must be **VALIDATED** by manufacturer / method developer
  - define / validate ***performance specification***

- Tests can only be **VERIFIED** against a validated ***performance specification***
  - CE marked kit
  - Internal technical validation



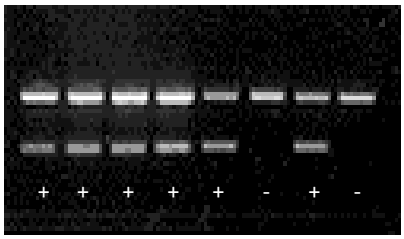
Performance specification

Should comprise (at least):

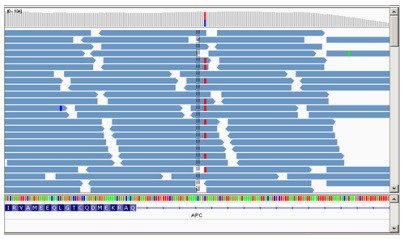
- An estimate of the test **ACCURACY** including measurement uncertainty (e.g. confidence limits)
- Limitations on **critical parameters** that will ensure the desired level of accuracy.
- Control measures required for monitoring routine maintenance of this level of accuracy

Types of tests

- All tests are fundamentally quantitative
- Sometimes we use the quantitative result directly
- However, it is often necessary to make an inference about the sample based on the quantitative result - ***Qualitative***

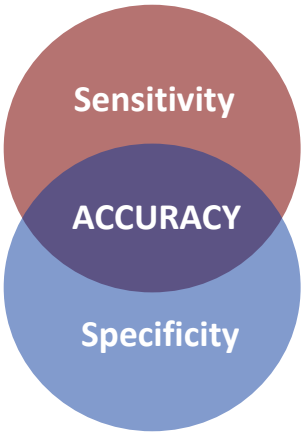


NM_000492.3(CFTR):c.1521_1523delCTT (p.Phe508delPhe)



NM_000038.5(APC):c.643C>T (p.Gln215Ter)

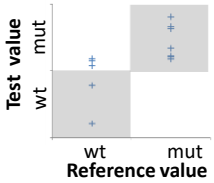
Components of Accuracy Qualitative tests:

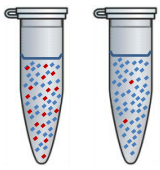


- describes how good the test is at detecting positives (Variant alleles)
- describes how good the test is at detecting negatives (Reference alleles)

Sensitivity

- … a measure of how good an assay is at detecting true positives (Qualitative)
- … a measure of how little of an analyte can be detected in a mixture (Quantitative - Limit of Detection)
- … a measure of responsiveness to changes in conditions





Determination of accuracy

- Assess the performance (**ACCURACY**) of the test in comparison with a 'gold standard'
- Gold standard is a set of control samples that have mutational status assigned without error (or best available)

2×2 contingency table

	Gold standard positive	Gold standard negative
Test Positive	True Positive (TP)	False Positive (FP)
Test Negative	False Negative (FN)	True Negative (TN)

2×2 contingency table

	Gold standard positive	Gold standard negative
Test Positive	True Positive (TP)	False Positive (FP)
Test Negative	False Negative (FN)	True Negative (TN)

Sensitivity = $\frac{\sum True\ Positive}{\sum Gold\ Standard\ Positive}$

2×2 contingency table

	Gold standard positive	Gold standard negative
Test Positive	True Positive (TP)	False Positive (FP)
Test Negative	False Negative (FN)	True Negative (TN)
	Sensitivity = $\frac{\sum True\ Positive}{\sum Gold\ Standard\ Positive}$	Specificity = $\frac{\sum True\ Negative}{\sum Gold\ Standard\ Negative}$

2×2 contingency table

	Gold standard positive	Gold standard negative	
Test Positive	True Positive (TP)	False Positive (FP)	Positive Predictive Value (PPV) = $\frac{\sum True\ Positive}{\sum Test\ Positive}$
Test Negative	False Negative (FN)	True Negative (TN)	
	Sensitivity = $\frac{\sum True\ Positive}{\sum Gold\ Standard\ Positive}$	Specificity = $\frac{\sum True\ Negative}{\sum Gold\ Standard\ Negative}$	

2×2 contingency table

	Gold standard positive	Gold standard negative	
Test Positive	True Positive (TP)	False Positive (FP)	Positive Predictive Value (PPV) = $\frac{\sum True\ Positive}{\sum Test\ Positive}$
Test Negative	False Negative (FN)	True Negative (TN)	Negative Predictive Value (NPV) = $\frac{\sum True\ Negative}{\sum Test\ Negative}$
	Sensitivity = $\frac{\sum True\ Positive}{\sum Gold\ Standard\ Positive}$	Specificity = $\frac{\sum True\ Negative}{\sum Gold\ Standard\ Negative}$	

Measurement: General considerations

- Analyses should be blinded
- Controls should not have been used to optimise the assay
- Requirements for negatives (specificity)
- Positives should:
  - be representative of what is expected in practice
  - reflect known strengths **and** weaknesses of the methodology

Sensitivity

	Sensitivity
More than 500 mutations have been identified in the CFTR gene, making it an excellent system for testing mutation scanning techniques. To assess the sensitivity of denaturing gradient gel electrophoresis (DGGE), we collected a representative group of 202 CFTR mutations. All mutations analyzed were detected by scanning methods other than the DGGE approach evaluated in this study. DGGE analysis was performed on 24 of the 27 exons and their flanking splice site sequences. **After optimization, 201 of the 202 control samples produced an altered migration pattern in the region in which an alteration occurred. The remaining sample was sequenced and found not to have the reported mutation**. The ability of DGGE to identify novel mutations was evaluated in three Asian CF patients with four unknown CF alleles. Three novel Asian mutations were detected-K166E, L568X, and 3121-2 A-->G (in homozygosity)-accounting for all CF alleles. **These results indicate that an optimized DGGE scanning strategy is highly sensitive and specific and can detect 100% of mutations.**	100%
...**A larger set of 32 mutant DNA specimens was then analyzed** using these optimized tandem CAE-SSCP/HA protocols and materials and **yielded 100% sensitivity of mutation detection**...	100%

Example

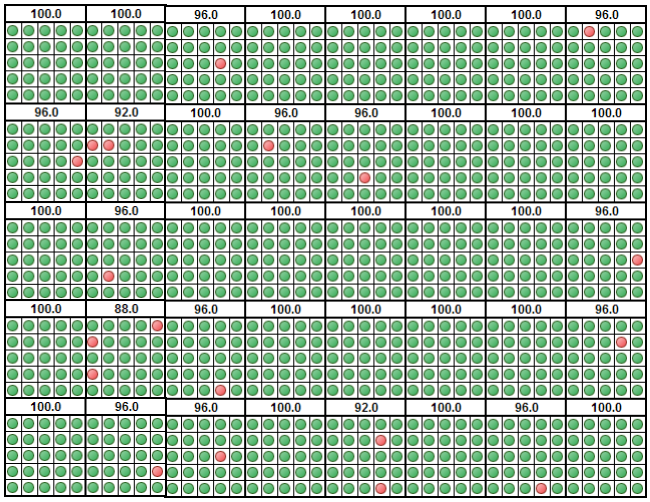
Test 1  Sensitivity = 25/25 = 100%

Test 2  Sensitivity = 23/25 = 92%

Test 3  Sensitivity = 24/25 = 96%

Test 4  Sensitivity = 22/25 = 88%

Example



Overall sensitivity = 984/1000 = 98.4%

Sampling Error

- ***Sampling error*** is incurred when the statistical characteristics of a population are estimated from a subset, or sample, of that population
- The larger the sample the smaller the error
- The 'worst-case' sensitivity should be calculated account for sampling error

Lower limit of 95% confidence interval is highest reportable sensitivity

Rule of Three

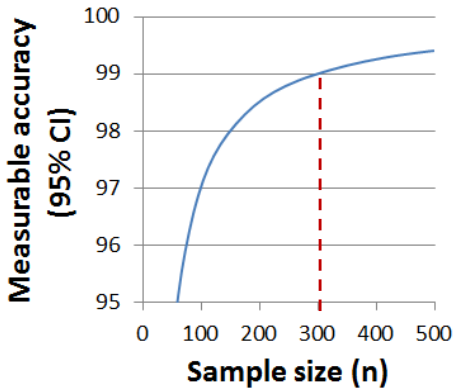
For the 95% confidence interval

The probability of NOT seeing a FN in a validation of sample size ***n*** ≈ 3/***n***

Max possible sensitivity is 1 ≈ 1 – 3/30
So for ***n*** =30 Max reportable sensitivity ≈ 1 – 0.1 = 0.9

This is true for any proportional value for ***n*** ≥ 20

Sample size calculation



Accuracy ≥ 99% (95% CI)

Calculation of sensitivity

MedCalc®
easy-to-use statistical software

http://www.medcalc.org/calc/diagnostic_test.php

	Disease				
Test	**Present**	**n**	**Absent**	**n**	**Total**
Positive	True Positive	a=298	False Positive	b=2	a + b = **300**
Negative	False Negative	c=2	True Negative	d=150	c + d = **152**
Total		a + c = **300**		b + d = **152**	

Test

Results

Sensitivity	$\frac{a}{a+c}$	= 99.33 %	95% CI: **97.61 %** to **99.90 %**
Specificity	$\frac{d}{b+d}$	= 98.68 %	95% CI: **95.32 %** to **99.80 %**

Sensitivity

	Sensitivity with 95% confidence
More than 500 mutations have been identified in the CFTR gene, making it an excellent system for testing mutation scanning techniques. To assess the sensitivity of denaturing gradient gel electrophoresis (DGGE), we collected a representative group of 202 CFTR mutations. All mutations analyzed were detected by scanning methods other than the DGGE approach evaluated in this study. DGGE analysis was performed on 24 of the 27 exons and their flanking splice site sequences. **After optimization, 201 of the 202 control samples produced an altered migration pattern in the region in which an alteration occurred. The remaining sample was sequenced and found not to have the reported mutation**. The ability of DGGE to identify novel mutations was evaluated in three Asian CF patients with four unknown CF alleles. Three novel Asian mutations were detected-K166E, L568X, and 3121-2 A-->G (in homozygosity)-accounting for all CF alleles. **These results indicate that an optimized DGGE scanning strategy is highly sensitive and specific and can detect 100% of mutations.**	≥98.5%
...**A larger set of 32 mutant DNA specimens was then analyzed** using these optimized tandem CAE-SSCP/HA protocols and materials and **yielded 100% sensitivity of mutation detection**...	≥91.6%

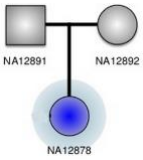
Example:
Technical validation of NGS based mutation scanning for constitutional variants

Strategy

- TruSight One Sequencing Panel (Illumina)
- 4,813 genes with known clinical significance
- Virtual in-silico panels for analysis

Research benefits
One validation covers all panels
Off-the-shelf availability
Cost
Single protocol
Equity of access
Flexibility
Future proofing

Design



Runs

- 3x NextSeq runs
- 3 x replicates of 4 samples on each run (12)

NIST RM 8398

Critical parameters

- Sensitivity
- Specificity
- Repeatability

} Minimum depth required

- Coverage robustness
- Robustness wrt sample input (extraction method)
- Analysis pipeline (indels)

Sensitivity

Genome in a Bottle
Authoritative Characterization of
Benchmark Human Genomes

- NA12878 high confidence variant calls (Genome in a bottle – GIAB)
  https://sites.stanford.edu/abms/giab
- >7000 variants in ROI
- Analysed separately: Se = TP / (TP + FN)
- Validation Se = Mean result

Technical validation: Sensitivity

MedCalc®
easy-to-use statistical software

Type of Variant	n	Sensitivity (95% Confidence interval)*		
Single nucleotide variants	7027		99.7 - 99.9	
Insertion and deletion variants (≤ 4 bp)	270	97.7 - 99.9	97.1 - 99.6	**99.7 -99.9**
Insertion and deletion variants (≥ 5 bp)	43	89.2 - 99.9		

- Exceeds minimum required standard of ACGS best practice guidelines
- Negligible variation in detection rate between analyses within run and between runs

Documentation for ISO accreditation is in draft

Key points

- Sensitivity is an ***estimate*** based on sub-sampling the population of all possible variants
  - A thorough validation will only examine a fraction of all possible variants
  - 20kb ROI has 100,000 possible SNV + 1bp indels
  - 300 positive controls = 0.3%
- Different regions perform differently ***but*** confidence range reflects this variation
- High numbers of variants = high confidence
  - >7000 variants - 95% confidence range <1%
- Technical variables that could affect detection should be evaluated in the validation to define limitations
  - Variant type (SNV, indel)
  - Context (GC content, repeats, homopolymers)
  - Run variability

Specificity

>20x

	A2657	A63A6	A639R
Total ROI count	62,309	62,309	62,309
ROI with no coverage	2,798	1107	883
ROI analysed	59,511	61,202	61,426
FP detected	1,452	1609	1334
Specificity	97.6%	97.4%	97.8%
95%CI	±0.1%	±0.1%	±0.1%

Interpretation	At 20x coverage overall specificity meets the required standard (i.e. >95% at 95%CI)
Outcome limitations	Coverage for performance specification is ≥20x All detected variant will be confirmed with a separate test

Key points

- ***estimate*** based on the assumption the gold standard test detects all variants
  - Publicly available data may not have been analysed equivalently
- Case is needed to determine an appropriate scale to measure specificity (i.e. the denominator)
  - Sample, Gene, ROI, Exon, per base?
- Level of criticality dependent on orthogonal confirmation
  - Correct result
  - TAT
  - Financial

Technical Verification

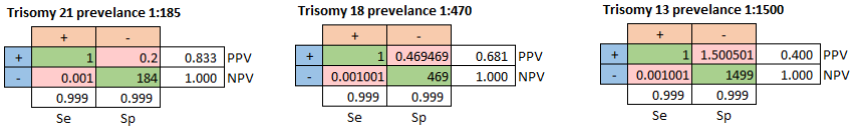
- 2 runs x 35 samples +1 control (72 analyses)
- Reproducibility of critical measurements across different samples and between runs
- Included 42 cases with known molecular diagnoses:
  - 41 detected, 1 in region with no read coverage)

Technology fully functional for true patient samples

Positive predictive value

- PPV measures how good a test is at predicting disease condition from a test result
- Proportion of positive results that are correct
- Considers population prevalence
- Particularly relevant to screening tests

NIPT for detection of aneuploidy



Genet Med. 2015 Mar;17(3):234-6. doi: 10.1038/gim.2014.92. Epub 2014 Aug 7.
Discordant noninvasive prenatal testing and cytogenetic results: a study of 109 consecutive cases.
Wang JC[1], Sahoo T[2], Schonberg S[3], Kopita KA[1], Ross L[1], Patek K[3], Strom CM[1].

EJHG Open

European Journal of Human Genetics (2010), 1–13

www.nature.com/ejhg

ARTICLE

A standardized framework for the validation and verification of clinical molecular genetic tests

Christopher J Mattocks*[,1,7], Michael A Morris[2,7], Gert Matthijs[3,7], Elfriede Swinnen[3], Anniek Corveleyn[3], Els Dequeker[3], Clemens R Müller[4], Victoria Pratt[5] and Andrew Wallace[6], for the EuroGentest Validation Group[8]

http://www.clsi.org/

CLINICAL AND LABORATORY STANDARDS INSTITUTE®