Statistician's Perspective of Multiplex Assays

Estelle Russek-Cohen, PhD
Division of Biostatistics
CBER/FDA

Disclaimer

- The presentation reflects the views of the author and are not necessarily the views of the FDA.
- Statisticians rely on clinical and scientific colleagues to determine what is or is not appropriate performance for a test. Asking a statistician to tell you the sample size without providing more info is……

Outline

- Multiplex Assays
- Statistics 101 for Diagnostic devices
- Genetic tests
- Multiplex pathogen tests
- Conclusions

Multiplex Assays

- “Two or more targets simultaneously detected via a common process of sample preparation, target or signal amplification, allele discrimination, and collective interpretation.”
- CLSI Guideline: MM17

  Verification and Validation of Multiplex Nucleic Acid Assays

Examples

- Human genotyping

  RBC Antigens (CBER) CYP450 Genes (CDRH)
- Multiple pathogen detection

  Blood donor screening (CBER)

  Respiratory virus panel (CDRH)
- Pathogen subtyping

  HIV Genotyping (CBER)

  Flu Genotyping (CDRH)

Clinical Reference Standard

- "considered to be the best available method for establishing the presence or absence of the target condition…it can be a single test or method, or a combination of methods and techniques, including clinical follow-up"
- does not consider outcome of new test under evaluation (see *discrepant resolution* in FDA guidance (2007))

Clinical Reference Standard: for “Target condition of interest”

- Presumes patients are from intended use population.
- Clinical reference standard is usually a review office decision

  Can combine clinical (e.g. symptoms)

  and analytical results
- Genotyping for inherited genes:

  Bidirectional sequencing with quality check

Some basics:
Statistics 101 for Diagnostic Devices

- One marker at a time compared with

  clinical reference standard:

Sensitivity=

P(test : detected| target condition "A" there)

Specificity=

P(test : not detected| target condition "A" not there)

- Need two values per "output"

Need 95% confidence intervals for sensitivity and specificity (See FDA Guidance (2007))

Statistical methods for agreement measures similar.

Estimates of "Sensitivity"/ Specificity Score(CDRH) or Exact(CBER)

Number specimens	Observed Performance	95% (2-sided) Lower Conf. Bound
5	5/5=100%	56.5% 47.8%
30	30/30=100%	88.6% 88.7%
35	35/35=100%	90.1%
60	60/60=100%	94.0%
120	120/120=100%	96.9%
10000	**10000/10000**	**99.96%**

Estimates of "PPA"/ "NPA"

Score(CDRH) or Exact(CBER)

Number specimens	Observed Performance	95% (2-sided) Lower Conf. Bound
5	5/5=100%	56.5% 47.8%
30	30/30=100%	88.6% 88.7%
35	35/35=100%	90.1%
60	60/60=100%	94.0%
120	120/120=100%	96.9%
10000	**10000/10000**	**99.96%**

Considering prevalence in prospective studies: Negative and Positive Predictive values

- Negative Predictive Value (NPV):

  $$\frac{\text{Num without condition of interest}}{\text{Number that test negative}}$$

- Positive Predictive value (PPV):

  $$\frac{\text{Num with condition of interest}}{\text{Number that test positive}}$$

- NPV and PPV depend on prevalence
- Assumes: clinical reference standard

Notation

	Interpretation via clinical reference std	
Device output	Target condition +	Target condition -
Target condition +	True Positive (TP)	False Positive (FP)
Target condition -	False Negative (FN)	True Negative (TN)

Prevalence (TP+TN)/(Total sample size)

Sensitivity =100*TP/(TP+FN); Specificity=100*TN/(TN+FP)

PPV= 100*TP/(TP+ FP); NPV=100*TN/(TN+FN)

Prevalence 9%

	Interpretation via clinical reference std	
Device output	Target condition +	Target condition -
Target condition +	99	10
Target condition -	1	990

Prevalence 100/1100= 9%
SE=99/100= 99% SP=990/1000=99%
PPV= 99/109= 91% NPV=990/991= 99.9%

Prevalence 1%

	Interpretation via clinical reference std	
Device output	Target condition +	Target condition -
Target condition +	99	100
Target condition -	1	9900

Prevalence 100/10100= 1%
SE=99/100= 99% SP=9900/10000=99%
PPV= 99/199= 50% NPV=9900/9901= 99.99%

Guidances/Guidelines

- CLSI MM17 Multiplex; EP5 Precision
- FDA:

Pharmacogenetic tests and genetic tests for heritable markers

Multiplex instrumentation

Reporting of Results for Diagnostic Tests

Special control guidances

Note: Blood donor pathogen testing: BLA!

Heritable Markers Guidance

- Some devices have multiple intended uses
- More than one study may be needed

  e.g. carriers or those with congenital disease
- Rare alleles, mutations, genotypes:

  Challenge
- Need supporting literature to justify reporting of specific mutations
- May still need big study. Role of banked specimens is review office decision.

Heritable Markers Guidance: Device Design

- All elements of sample prep need to be documented
- Matrix type(s) specified
- Test platform and technology, buffers, etc
- How much DNA and/or sample is needed to run the test? What proof do you have?
- Internal and external controls?

What **statisticians** need to know

- What does your device report
- Which genotypes/SNPs/Alleles does it detect
- What algorithms does it use to interpret results…was it clearly defined before the start of the study?

  e.g. how were cutoffs established?

  how many of each genotype, etc are you anticipating in your study?
- Are there ambiguous results?

Factors likely to influence study size

- Number of outputs per gene
  - Relevance of Haplotype versus SNP
- Number of genes
- Rareness of some results
  - Appropriateness of enriched studies?
- Heterogeneity across ethnic groups
  - Capture diverse study sites
- Clinical consequences of mistakes

Blood RBC Genotyping

- Donors and patients that are mismatched: Patients can die or build up antibodies faster
- There are 29 systems for typing red blood cells:

  ABO Rh+/- ….. Kell JMH

Cannot rely solely on a genotyping test:

Patients may receive blood from multiple donors at once or may need repeat transfusions

Blood often divided into component parts:

More patients impacted by errors

Interpreting RBC Genotyping assays

- One SNP at a time?
- Phenotype?

  Serology is better understood
- Overall concordance is insufficient:

  one number does not capture performance
- Need: Probability “device reports genotype mapping to correct result for serotype class” given “patient’s serotype”

Interpreting the Assays

- For common phenotypes:

  95% Lower confidence bound >99%
- Rare genotypes or phenotypes: talk to CBER
- If only comparing to traditional serology, can only claim agreement on the label
- Also report agreement to serology on patient level
- Expect to provide supporting literature.

Precision studies

- Panel members: with true genotype known

  Common genotypes

  Genotypes that challenge the device

  External controls
- Capture all assay steps (e.g. DNA extraction) (MM17)
- 4 sites (3 external and one internal)
- Sources:

sites, operators, instruments, day to day and lot to lot

role of “confounding”: good and bad

Multiplex pathogen assays

- Blood donor screening

  If you are going to test 5 pathogens..

  convenience and efficiency of single device
- Concerns over devices that test for lots of bugs:

  How many false positives? (Alarm fatigue?)

  Matrix may not be optimized for all
- Recognizing co-infections (see MM17)

Analytic Studies (CBER)

- Limit of detection (probit analysis..3 lots)

  Calculated for WHO standard

  Validated for other genotypes
- Levels for precision studies: driven by LOD.
- Analytic specificity:

  will other pathogens mess up my test?
- Endogenous and Exogenous interferents

  Exogenous interferents: fewer in healthy pop?

Conclusions

Be realistic when designing the device

Consider doing HW on frequency of various genotypes or pathogens your device will detect.. does this vary by site?

Submit a pre-IDE including analytical and clinical studies.

For genetic markers or devices that capture lots of pathogens:

Unlikely: a simple formula for sample size

References

- Pepe (2003) The Statistical Evaluation of Medical Tests for Classification and Prediction Oxford Press
- Zhou et al (2002) Statistical Methods in Diagnostic Medicine Wiley

More references

- Instrumentation for Clinical Multiplex Test Systems: Guidance for Industry and FDA Staff

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm077819.htm 2005

- Pharmacogenetic tests and genetic tests for heritable markers June 2007

- FDA posts guidances on public website.

Still More references

- FDA(2007) Guidance for Industry and FDA Staff: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. March 2007
- FDA (2008) Draft Guidance: Establishing Performance Characteristics of In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Influenza Viruses. February 2008
- FDA (2009) Guidance for Industry and FDA Staff: Class II Special Controls Guidance Documents: Respiratory Virus Panel Multiplex Nucleic Acid Assay October 2009

Slides not presented

Other Considerations for Precision Studies for "Heritable Genotype" Devices

- Specimen type, amt of specimen per result and stability (5 days?)
- Numbers:

  measurements from single sample

  results per run; results per day
- Logistics of switching operators; switching lots, etc.
- EP5 principles useful but EP5 more focused on plasma and serum sample assays.
- Can assess accuracy for each panel member