AACC *and the Management Sciences and Patient Safety Division Present*

Statistical Methods for Clinical Laboratorians

Method Validation and Reference Ranges : Part 2

CLSI Precision Protocol

- EP5-A2
  - Guidelines intended primarily for manufacturers of *in vitro* diagnostic methods
- EP15-A2
  - Guidelines for validation of manufacturer's method performance specifications
- Terminology
  - Repeatability vs. reproducibility (ISO standards)
  - Sample vs. specimen
  - Total precision vs. within-laboratory precision

CLSI Guideline EP5-A2

- Guidelines for manufacturers of IVD devices and methods (or user-modified methods)
- Four elements define conditions: time, operator, calibration, instrument
- Familiarization period (5 days)
- Day-to-day precision
  - Two concentrations twice a day for 20 days
- Within-run precision
  - 20 consecutive assays
  - Recommend “pooling” results of several within-run evaluations
- Outliers (> 5.5 SD)

CLSI Guideline EP15-A2

- Specifies how a laboratory should verify the analytical precision claims made by the manufacturer
- Procedure
  - Two specimens once a day in triplicate, for five consecutive days
  - Triplicates define within-run precision
- If the SD (or CV) is less than the manufacturer's specifications, then the precision claim is validated
- If the SD (or CV) is greater than the manufacturer's specifications, then a Chi-square test can be used to assess the statistical significance of the difference

Validation of Reference Intervals

- CLIA 88 requires validation of reference intervals
  - 42 CFR 493.1253 (Subpart K):

    Verify that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population.

  - CLIA does not specify how reference intervals are to be validated
  - Note that CLIA does not require laboratories to establish their own reference intervals, only to verify that manufacturer's reference intervals are appropriate for their patient population
- CLSI has provided guidance in meeting this CLIA standard

CLSI Document C28-A3

- "Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition" (2008)
- Transference (adoption of a reference interval established by another laboratory)
  - Comparability of analytical systems
  - Comparability of test subject population
- Validation
  - Subjective method (non-statistical)
  - Small sample method
  - Larger sample method

Comparability of Analytical Methods

- Are they the same platform and method?
  - Some platforms support more than one method
  - Some methods can be adapted to multiple platforms
- Is the measurement principle the same?
  - Nephelometry vs. chemiluminescent immunoassay
  - "One-step" vs. "two-step" methods (*e.g.*, free T4)
  - Direct vs. indirect potentiometry
- Do the methods correlate with each other?
  - Slope and y-intercept; standard error of the estimate
- If a proportional bias exists between the methods (slope $\neq$ 1.0), can the reference range be adjusted based on the regression equation?

Comparability of Test Subject Populations

- Demographic factors that may influence the reference intervals for laboratory tests:
  - Age (pediatric vs. elder care)
  - Sex (women's hospitals)
  - Race (relatively minor)
  - Altitude (hematology)
  - Latitude (Vitamin D)
  - Diet
- Equivalence of the analytical methods and the test subject populations are the basis of the *subjective* method of reference interval transference

Small Sample Validation

- Procedure:
  - Select 20 healthy subjects in the local population
  - If at least 18 of the 20 fall within the reference interval to be transferred, the range is validated
  - If 16 or 17 fall within the reference interval, select 20 additional subjects, and if at least 18 of the second set of subjects fall within the reference interval, the range is validated
  - If neither of the above conditions is met, the range is not validated
- The statistical basis for this validation test is a binomial distribution
  - $r > 2$; $n = 20$; $p = 0.05$; $P(>2;20;0.05) = 0.075$

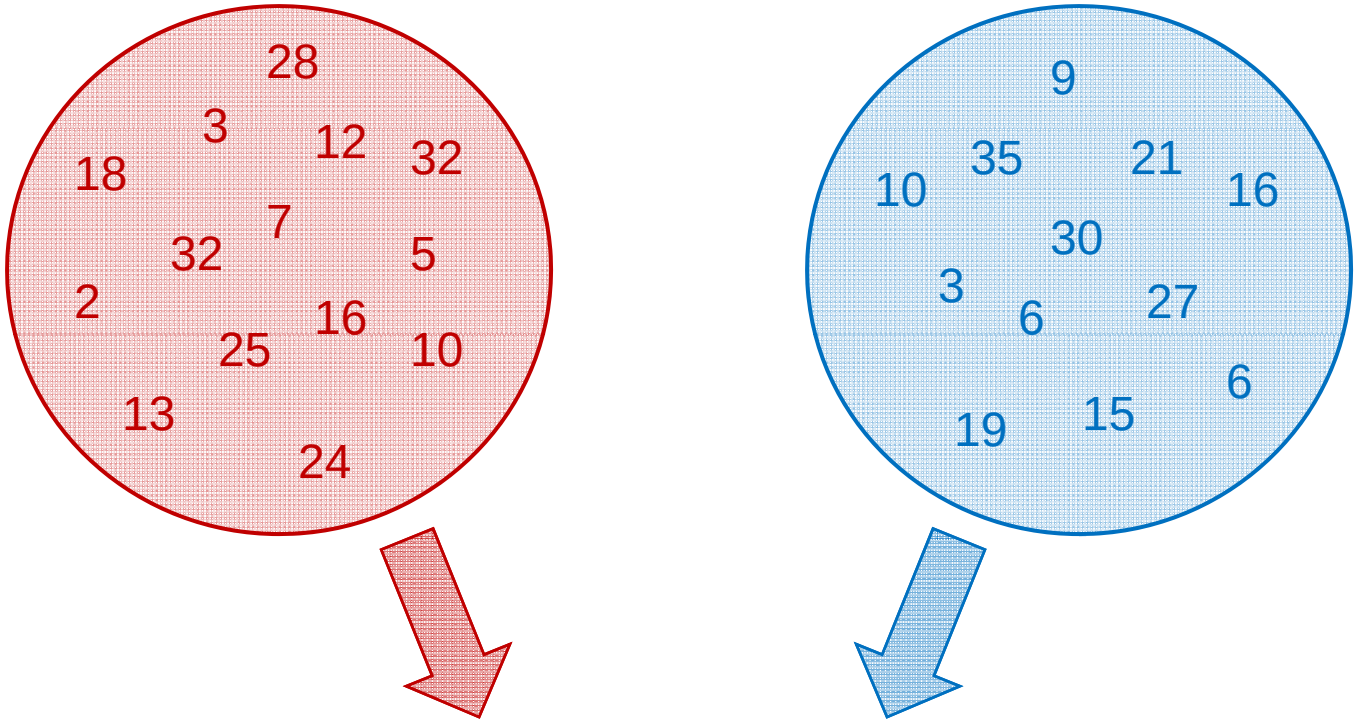
Larger Sample Validation

- Procedure
  - Select 60 (or more, but less than 120) healthy individuals as the local reference population
  - Calculate the reference range (non-parametrically) for the local reference population
  - Compare the calculated reference range for the local reference population with the range to be transferred
- Alternatively, the local and reference data can be compared by a robust (non-parametric) method

Parametric vs. Non-Parametric Statistics

- Parametric methods are based on statistical distributions
  - Examples: Student's *t*, χ^2, *F*-test, *etc.*
  - Statistical distributions characterize data with a limited set of descriptors, such as mean and standard deviation
- Non-parametric methods do not assume any particular distribution of the data
  - Sometimes called "distribution-free methods"
  - Generally requires raw data, rather than mean and standard deviation
  - Examples: Mann-Whitney U (or Wilcoxen rank-sum) test, Spearman's rank correlation coefficient, Wald–Wolfowitz runs test
- Reference ranges often have irregular distributions

Mann-Whitney U Test



2,3,3,5,6,6,7,9,10,10,12,13,15,16,16,18,19,21,24,25,27,28,30,32,32,35

Mann-Whitney *P*=0.48

Monte Carlo Simulations

Reference database

$$\bar{x}_{\bar{x}} \pm SD_{\bar{x}}$$

N = 20
Mean, SD

N = 20
Mean, SD

N = 20
Mean, SD

Key Concepts

- Method “sensitivity” is an ambiguous concept with multiple definitions, and should be replaced with the concept of “detection limit”
- Forensic drug testing laboratories observe strict legal definitions of “limit of detection” and “limit of quantitation”
- The CLSI has recommended protocols for assessing the precision of clinical laboratory methods both for vendors and for end users
- There are various methods for validating reference ranges, from simple comparison to non-parametric statistical methods

1850 K Street, NW, Suite 625
Washington, DC 20006-2213
(800) 892-1400 or (202) 857-0717
www.aacc.org