Forum for Collaborative HIV Research External Validation of CD4 and Viral Load Assays

Paris, France
June 29, 2007

Thomas J. Spira, M.D.
International Laboratory Branch
Global AIDS Program
Centers for Disease Control & Prevention
Atlanta, GA, USA

CLIA Final Rule

- Non-Waived Tests Approved by FDA
  - *Verification of performance specifications.* Each laboratory that introduces an unmodified, FDA-cleared or approved test system must do the following before reporting patient test results:
    - Demonstrate that it can obtain performance specifications comparable to those established by the manufacturer for the following performance characteristics:
      - Accuracy
      - Precision
      - Reportable range of test results for the test system
    - Verify that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population

CLIA Final Rule

- This would generally mean performing four experiments:
  - A comparison of methods experiment to estimate inaccuracy or bias
  - A replication experiment to estimate imprecision
  - A linearity type of experiment to determinate the reportable range
  - Collect reference values to verify the reference range [alternatively, the laboratory's medical director can document that the manufacturer's ranges or textbook ranges are appropriate for the clientele being served]

Experiments for Estimating Analytic Errors

Type of Analytic Error	Evaluation Experiment	
	Preliminary	Final
Random Error	Replication Within Run	Replication Between Runs
Constant Error	Interference	Comparison Of Methods
Proportional Error	Recovery	

Westgard, JO, Basic Method Validation. 2nd Ed. Madison, Westgard QC Inc., 2003, p.51.

Experiments for Estimating Analytic Errors

- Replication experiment.
  - Provides information about random error
  - Performed by making measurements on a series of aliquots of the same test sample within a specific period of time, usually within an analytical run, within a day, or over a period of a month.
  - Preliminary experiment usually involves determining within-run imprecision.
  - Final experiment generally requires at least 20 working days to provide a good estimate of total imprecision, which includes within and between run components

Experiments for Estimating Analytic Errors

- Interference experiment
  - Provides information about the constant systematic error caused by the lack of specificity of the method
  - One test sample is prepared by adding the suspected material to a sample containing the analyte
  - A second sample of the original sample is diluted by the same amount with solvent, then both samples are analyzed by the test method and the difference determined

Experiments for Estimating Analytic Errors

- Recovery experiment
  - Provides information about the proportional systematic errors caused by a competitive reaction
  - A test sample is prepared by adding a standard solution of the analyte being tested to an aliquot of a patient specimen
  - A baseline sample is prepared by adding an equal amount of the solvent used for the standard solution to a second aliquot of the same patient specimen
  - The two samples are then analyzed by the test method and the amount recovered is compared to the amount added.

Experiments for Estimating Analytic Errors

- Comparison of methods experiment
  - Estimate the average systemic error observed with real patient samples, but can also reveal the constant or proportional nature of the error
  - A series of patient specimens are collected and analyzed by both the test method and a comparative analytic method
  - The results are compared to determine the differences between the methods, which are the analytical errors between the methods

Performance Characteristics
FDA 510 K Submission

Analytical Performance	FlowCare CD4	Guava EZCD4
Sites	4	4
Precision, Reproducibility	Commercially available hematology (Streck STaK-Chex, tri-level-low, normal, high) for WBC and lymphocyte count and percentage. CD4 controls (Streck CD-Chex Plus, bi-level-low, normal) analyzed on two instruments, in triplicate runs, three times per day over three days. Performed at company.	10 replicate whole blood specimens from each of 3 abnormal donors representing each of three CD4+ absolute count ranges (0-200, 201-500, 501-2000)

Performance Characteristics
FDA 510 K Submission

Analytical Performance	FlowCare CD4	Guava EZCD4
Linearity/assay reportable range	Full range: Concentrated whole blood diluted autologous platelet poor plasma to achieve the desired concentration levels. The 50% (normal range) sample was used to determine the expected values at the other concentration levels. Low range: Same as above using blood from a donor with a CD4 count of approximately 400 cells/μL. The undiluted whole blood sample was used to determine the expected values at the other concentration levels.	Expected vs. Observed values of absolute CD4 T-cell counts on a preparation of a series of blood aliquots, each aliquot consisting of a decreasing volume of a bulk blood sample of known "high range absolute CD4+ T-cell count and an increasing volume of a bulk blood sample of known "low range" absolute CD4+ T-cell counts. All cell aliquots were prepared in duplicate and a total of 22 aliquots (11 pairs) were prepared.

Performance Characteristics FDA 510 K Submission

Analytical Performance (cont.)	**FlowCare CD4**	**Guava EZCD4**
Traceability, Stability, Expected values (controls, calibrators, or methods)	Antigen specificity of the CD4 monoclonal antibody submitted. Single use reagent tube.	Open vial stability testing
Detection limit	Not applicable.	Special indications were included in the linearity protocol for the concentration of the "low range" (<50) and "high range" (>2000) absolute CD4+ T-cell count bulk blood samples.
Comparison Studies		
Method Comparison with Predicate Device	Multi-site at 4 sites in US and sub-Saharan Africa using 403 normal (US) and abnormal (Africa) whole blood samples	Total of 365 abnormal donors in three absolute CD4+ T-cell count ranges were collected. Approximately 30 within each of the strata at each site.

FlowCare CD4 Linearity Study

Full Range Linearity:

Regression Statistics:	S/N 1816			S/N 2185		
N	WBC#	Lymph#	CD4#	WBC#	Lymph#	CD4#
	21	21	21	22	19	20
Correlation	0.998	0.967	0.981	0.984	0.994	0.992
Slope	1.018	0.960	1.002	0.957	0.969	0.950
Intercept	-131.732	-29.898	-85.463	204.348	-9.586	-7.461

Low Range Linearity:

Regressions Statistics:	CD4 Count	
	S/N 1816	S/N 1858
N	27	27
Correlation	0.994	0.994
Slope	1.042	0.975
Intercept	-12.932	13.295

FlowCare CD4 Linearity Study

FlowCare System Reportable Ranges:

Parameter	Reportable Range	Units
WBC Count	1.0 - 23.0	$10^3/\mu L$
Lymphocyte%	10 - 75	%
Lymphocyte Count	0.3 - 6.0	$10^3/\mu L$
CD4 Count	50 - 3000	µL
CD4%	0 – 80	%

FlowCare CD4 Comparison Study

Parameter	N	Corr. Coeff.	Slope	Intercept	Mean		Range	
					FlowCare	Reference	FlowCare	Reference
All Sites								
WBC#	403	0.9623	1.116	-335.13	6194.5	5851.9	1900-19700	1820-18100
LYM#	403	0.9145	1.081	-44.62	1956.8	1850.1	400-6500	500-5320
LYM%	403	0.9487	0.955	1.405	33.1	33.2	4.5-80.2	6.2-78.8
CD4#	425	0.9121	1.034	9.289	678.5	647.1	0-2318	7-2090
CD4%	401	0.9507	0.950	1.406	35.9	36.3	0.4-73.9	1.0-68.0

FlowCare CD4 Reference Range

FlowCare Parameter	N	Range		Mean ± SD
		Minimum	Maximum	
WBC Count	206	4400	12100	6882.5 ± 1618.76
Lymph Count	206	1100	3300	1963.6 ± 525.62
Lymph %	206	16.4	45.3	28.8 ± 6.21
CD4 Count	207	468	1702	928.5 ± 284.53
CD4%	205	31.3	65.8	47.7 ± 7.46

Whole blood specimens from apparently healthy males and females in the NE US, without selection on the basis of age or race. Expected results for the FlowCare parameter are presented based on a 95% normal distribution and compare closely with results observed with reference methods.

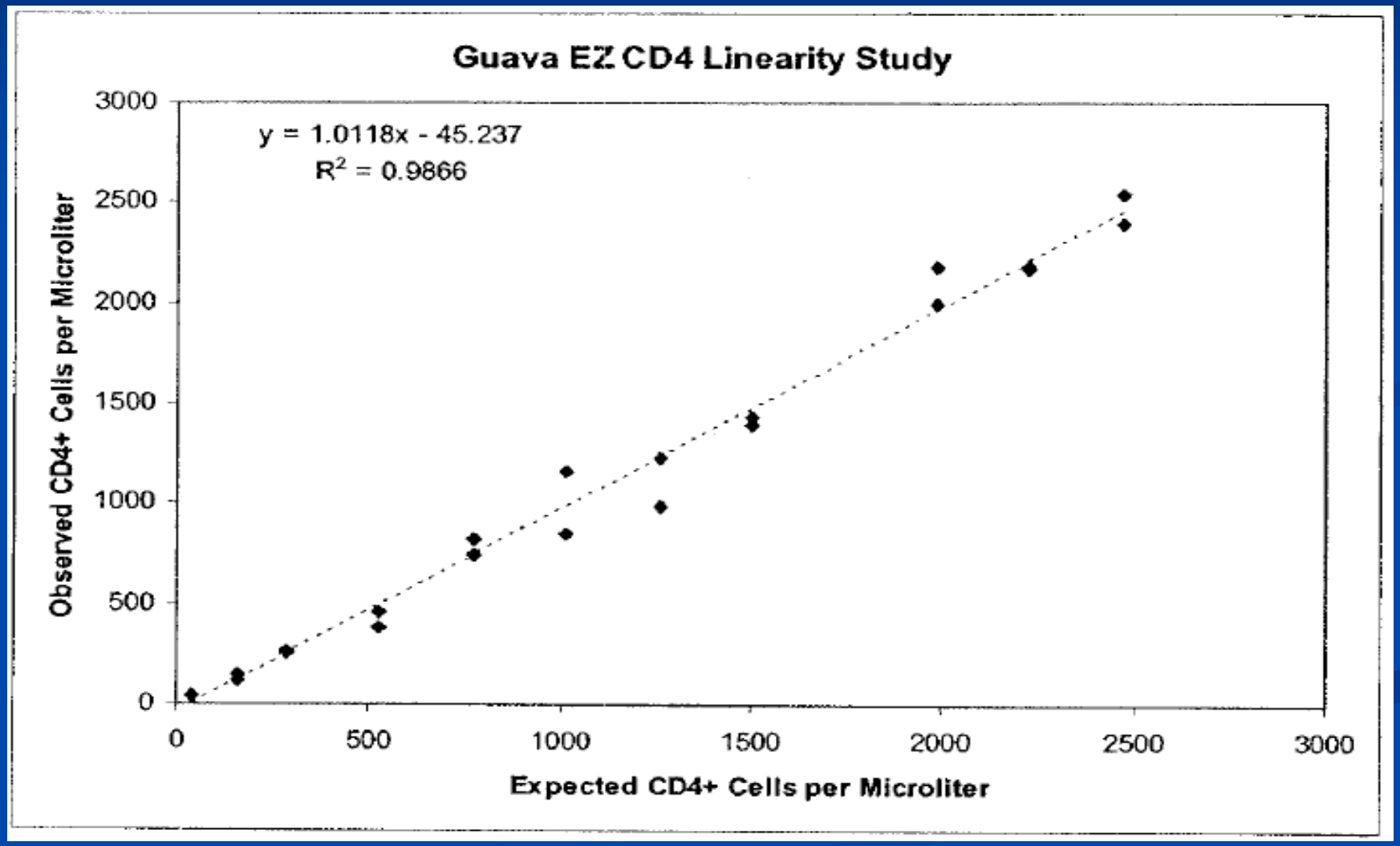
Guava EZCD4
Intra-laboratory Reproducibility

Study Site	Range	Mean EZCD4 CD4+ T Cells/µL	SD	CV	n
1	Low	178.41	24.08	13.50	10
	Mid	494.97	39.89	8.06	10
	High	676.45	32.16	4.75	10
2	Low	72.61	7.59	10.45	10
	Mid	417.27	30.06	7.20	10
	High	655.72	32.29	4.92	10
3	Low	81.82	9.04	11.05	10
	Mid	366.36	14.17	3.87	10
	High	870.43	30.99	3.56	10
4	Low	165.44	7.76	4.69	10
	Mid	373.20	13.79	3.69	10
	High	559.65	18.07	3.23	10

Guava EZCD4 Comparison Study

Study Site	n	R squared	Slope	Intercept	Range
1	92	0.95	+ 1.00	18.64	13-1465
2	91	0.93	+ 0.96	35.51	17-1175
3	88	0.98	+ 1.17	18.46	47-1439
4	94	0.98	+ 0.95	13.29	8-1076

Guava EZCD4 Linearity Study



Guava EZCD4 Carryover Study

Sample Numbers	Blood Sample Types	Replicate No. 1 (Absolute CD4 Counts)	Replicate No. 2 (Absolute CD4 Counts)	Replicate No. 3 (Absolute CD4 Counts)
Low 1-3	Low Range Sample	122.67	139.63	133.01
High 1-3	High Range Sample	672.40	765.43	655.53
Low 4-6	Low Range Sample	138.89	113.87	129.49
High 4-6	High Range Sample	684.91	745.29	763.55
Low 7-9	Low Range Sample	144.24	156.11	155.90
High 7-9	High Range Sample	766.90	725.13	784.11
Low 10-12	Low Range Sample	139.64	171.87	146.28
High 10-12	High Range Sample	812.82	732.84	915.32
Low 13-15	Low Range Sample	150.52	150.12	132.84
High 13-15	High Range Sample	734.46	752.94	817.92
Low 16-18	Low Range Sample	138.63	156.83	140.73
High 16-18	High Range Sample	790.88	767.51	805.09
Low 19-21	Low Range Sample	147.68	149.98	136.08

Statistical Comparison of Pre-High and Post-High Low Range Samples

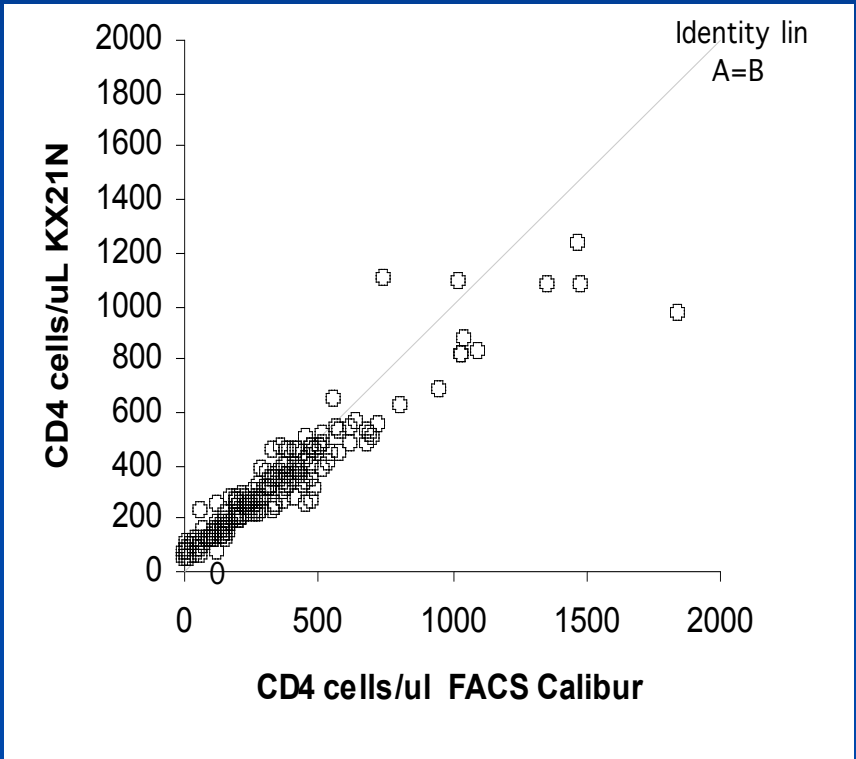
Low Range Sample Groups	Mean (Absolute CD4 Counts)	SD	CV (%)	n
Pre-High	139.19	9.26	6.65	7
Post-High	143.27	5.04	3.52	6

Sysmex K21N-Dynal Dynabeads Comparison Study – Abs. CD4 Methodology

	Methods to assess performance	Specimen Source	Sample Size	Distribution	CD4 T cell Strata
Part II	Acceptability Intra-assay variation (10 replicates)	Local	2 QC	1 Low 1 Mid	100-300 301-600
Part III-A	Run to run (10 replicates)	Local	2 QC	1 Low 1 Mid	100-300 301-600
Part III-B	Carry-over MHLMMLLHHM	Local	3	1 Low 1 Mid 1 Hi	100-300 301-600 >600
Part III-C	Inter assays (<6, 24, 48 hr)	Local	100	70 Low 30 Mid	100-300 301-600
Part III-D	Inter laboratory	Central	300	200 Low 100 Mid	100-300 301-600

Zimbabwe

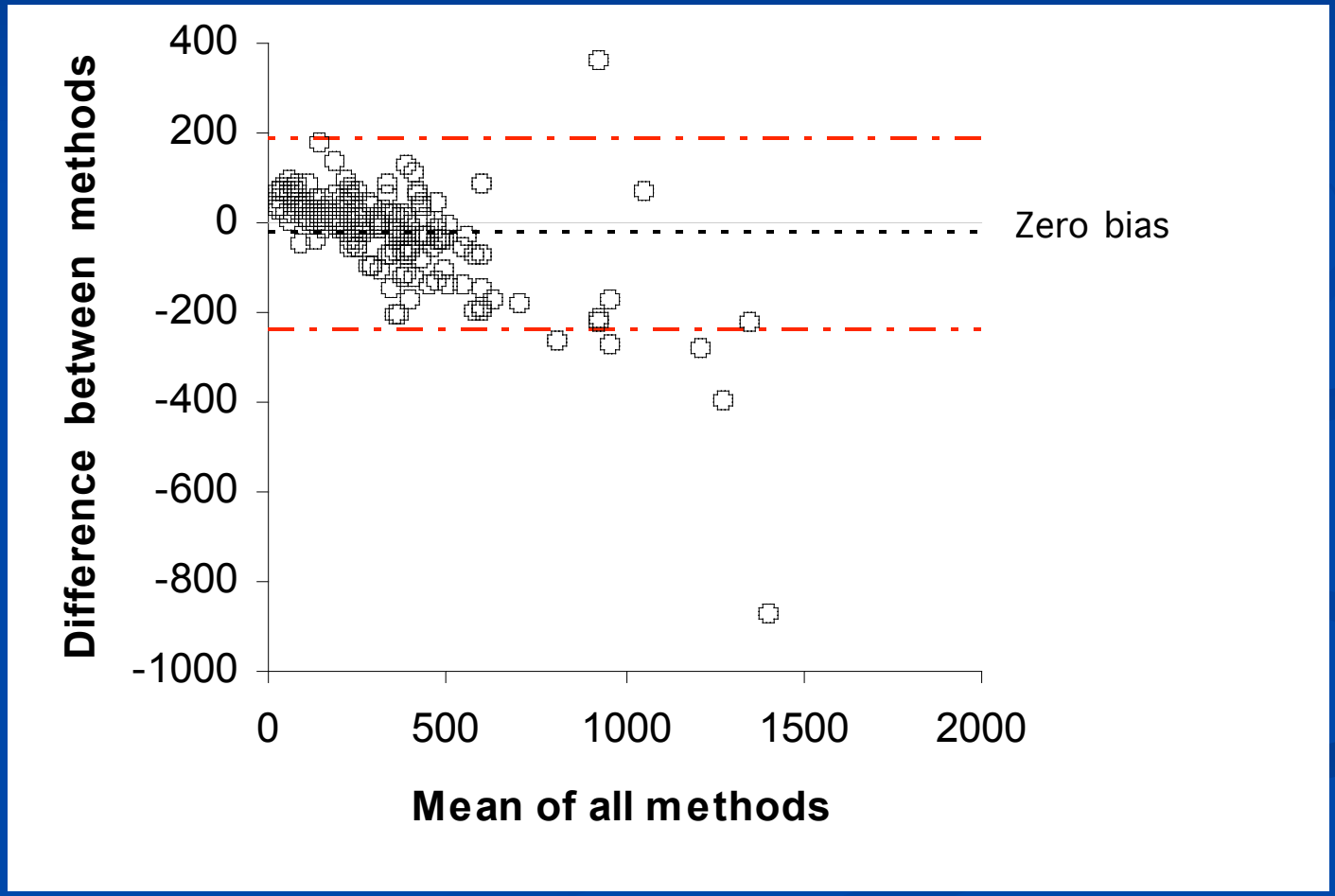
Sysmex K21N-Dynal Dynabeads Comparison Study – Abs. CD4 Regression Plot



n	181		
Bias	-24.569		
95% CI	-40.802	to -8.336	
95% limits of agreement		**95% CI**	
Lower	-241.489	-269.045	to -213.934
Upper	192.351	164.796	to 219.906

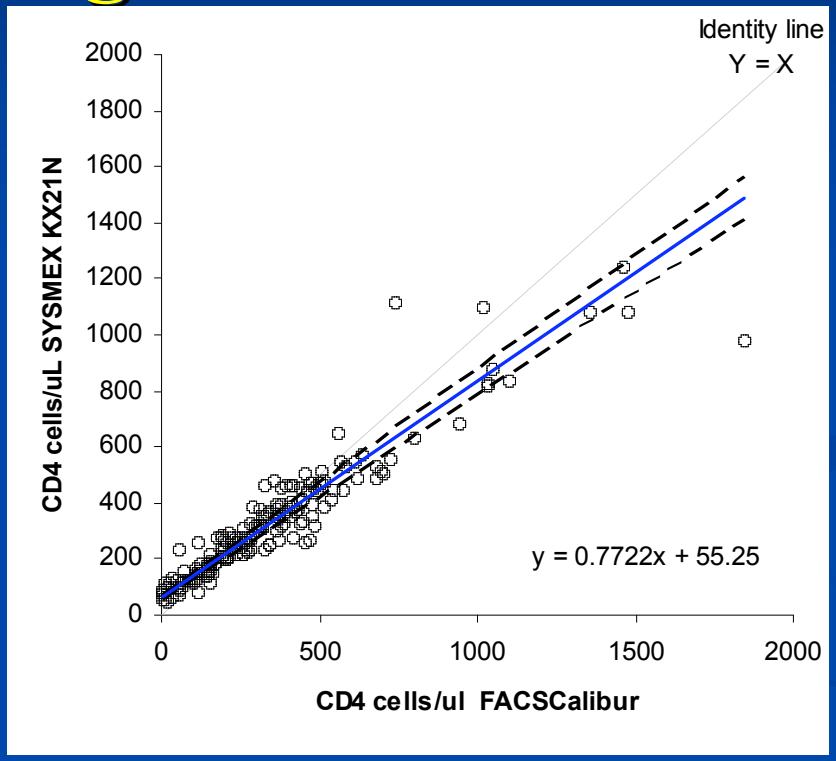
Zimbabwe

Sysmex K21N-Dynal Dynabeads Comparison Study – Abs. CD4 Bias Plot



Zimbabwe

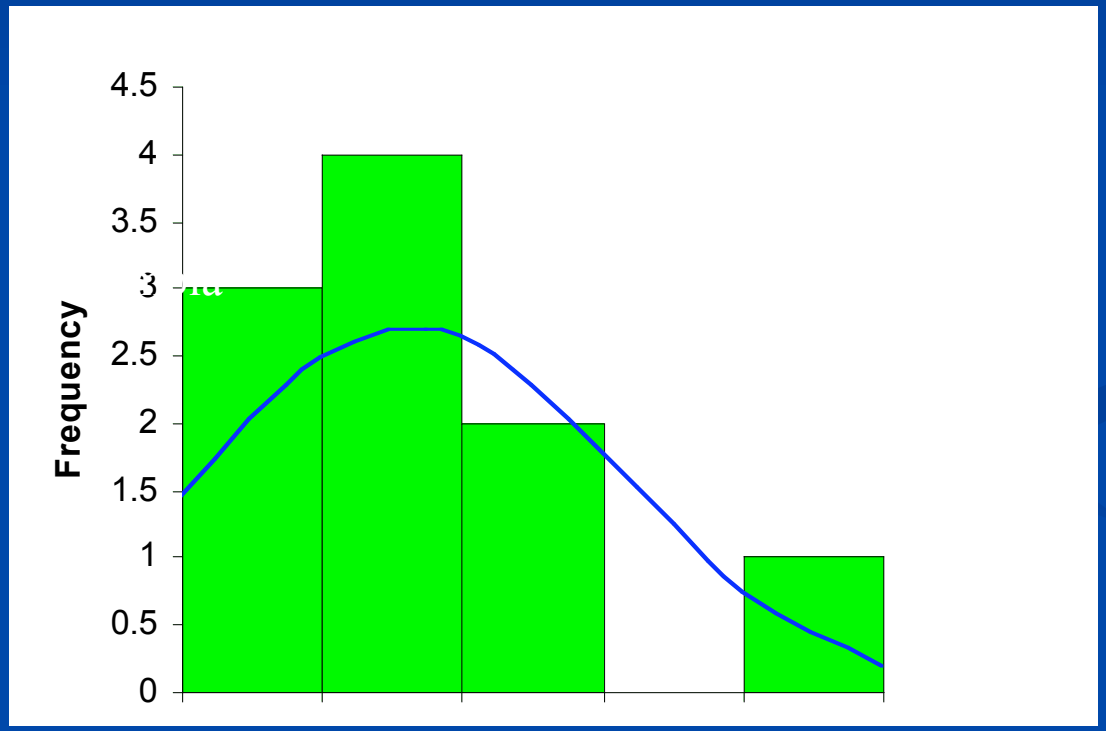
Sysmex K21N-Dynal Dynabeads Comparison Study – Abs. CD4 Passing and Bablok Method



n	181			
	Coefficient	95% CI		
Intercept	55.250	46.626	to 64.717	(constant bias detected)
Slope	0.772	0.734	to 0.810	(proportional bias detected)

Zimbabwe

Sysmex K21N-Dynal Dynabeads Reproducibility Study – Abs. CD4



n	10	
Mean	113.200	
95% CI	107.967	to 118.433
Variance	53.5111	
SD	7.3151	
SE	2.3132	
CV	6%	

Zimbabwe

Sysmex K21N-Dynal Dynabeads CD4 counts over 3 time periods

Evaluation of Compact Flow Cytometers and PLG Technique - Cote d'Ivoire

Accuracy Study	Two commercial stablized whole blood specimens of know target value and confidence interval (BD Multi-Check and BC Immunotrol.
Comparison Study	150 whole blood samples stratified with 50 HIV+ with CD4 count <200, 50 HIV+ with count between 200 and 499, and 50 with count ≥ 500 (30 HIV+, 20 HIV- FACSCalibur, EPICS XL, FACSCount, Guava, CyFlow Counter
Reproducibility Study	9 whole blood samples tested 5 times (3 samples from each CD4 stratum
Aged Specimens	15 samples held either stained or unstained for 24 and 48 hours prior to testing

Cote d'Ivoire Evaluation
Regression Analysis

Comparison (vs. FACSCalibur)	R (Spearman)	Y intercept
EPICS XL	0.99	-5.7
<200	0.98	-.004
200-499	0.94	-11.0
≥500	0.96	0.4
FACSCount	0.97	15.7
<200	0.92	2.6
200-499	0.92	-42.2
≥500	0.60	179.5
CyFlow Counter	0.98	-1.1
<200	0.94	7.7
200-499	0.88	-48.6
≥500	0.61	109.3
Guava	0.96	19.9
<200	0.92	11.1
200-499	0.64	30.2
≥500	0.62	236.8

Cote d'Ivoire Evaluation
Bias Analysis

Comparison (vs. FACSCalibur)	Mean	Range (± 1.96 SD)
EPICS XL	-3.6	-71.6-64.5
<200	0.1	-25.4-25.6
200-499	6.7	-33.7-47.2
≥500	-15.7	-117.3-85.8
FACSCount	17.5	-258.4-293.4
<200	7.1	-63.5-77.7
200-499	28.3	-67.3-123.9
≥500	19.2	-433.6-472
CyFlow Counter	23.4	-226.2-273.0
<200	11.5	-29.5-52.5
200-499	39.9	-31.6-111.4
≥500	22.1	-392-1-436.2
Guava	17.5	-261.6-296.6
<200	0.8	-54.8-56.3
200-499	50.3	-75.9-176.5
≥500	7.7	-442-4-457.7

Questions for Consideration

- Is there consensus on acceptable levels of performance for CD4 testing?
- Who should conduct validation studies of new technologies?
- Should each country do their own validation study?
- Are validation studies done at tertiary level labs adequate?
- Should validation protocols be standardized?
- Is multi-site, multi-country validation desirable?

Clinical Quality Requirements
Hematology Parameters

Clinical Quality Requirements

Test	Decision Level	D_{int}(%)	Source	s_{wsub}(%)	Source
Hematocrit	42 mg/dL	11.9%	Skendzel	2.5%	Fraser
Hemoglobin	15.0 g/dL	8.0%	Skendzel	2.4%	Fraser
Leukocyte count	$5x10^9$ cell/L $25x10^9$ cell/L	32% 28%	Skendzel Skendzel	15.6%	Fraser

Decision Intervals (Dint) expressed as a percentage change at a certain Decision Level (Dint = change divided by decision level multiplied by 100 to give a percentage).

Within-subject biological variation (s_{wsub})

Skendzel LP, Barnett RN, Platt R. Medically useful criteria for analytic performance of laboratory tests. Am J Clin Pathol 1985;83:200-205.
Fraser CG. Desirable standards for hematology tests: a proposal. Am J Clin Pathol 1987; 88:667-669.

WHO Draft Protocol
Protocol Design for Assessment of New Technologies for CD4 T-cell Enumeration

- Intra-laboratory variation
  - Sample size = 1 (local specimen)
  - 10 replicates
  - CD4 T-cell stratum = 100-300
  - Calculate mean, SD, CV%
- Inter-assay
  - Sample size = 7 CD4 T-cell stratum = 100-300
  - Sample size = 3 CD4 T-cell stratum = 301-550
  - Prepare samples within 6, 24, and 48 hours post draw
  - Calculate mean, SD, CV%

WHO Draft Protocol
Protocol Design for Assessment of New Technologies for CD4 T-cell Enumeration

- Instrument precision
  - Run-to-run
    - Sample size = 1
    - 10 runs
    - CD4 stratum = 100-300
    - Calculate mean, SD, CV%
  - Carryover
    - Sample size = 3
      - Low CD4 T-cell stratum = 100-300
      - Mid CD4 T-cell stratum = 301-600
      - High CD4 T-cell stratum = >600
      - Run the samples following this sequence HHHLLLHHHMMMLLL

WHO Draft Protocol
Protocol Design for Assessment of New Technologies for CD4 T-cell Enumeration

- Inter-laboratory variation-Reproducibility
  - Sample size = 25 CD4 T-cell stratum = 301-600
  - Sample size = 75 CD4 T-cell stratum = 100-300
  - Calculate mean, SD, and CV%
  - Number of participating sites = 3
  - Send-out specimens
  - All specimens are analyzed by the reference method and the test method to measure agreement
  - The performance of each lab has to be demonstrated prior to the start of the study by implementing intra-lab assay evaluation using both send-out samples and QC materials