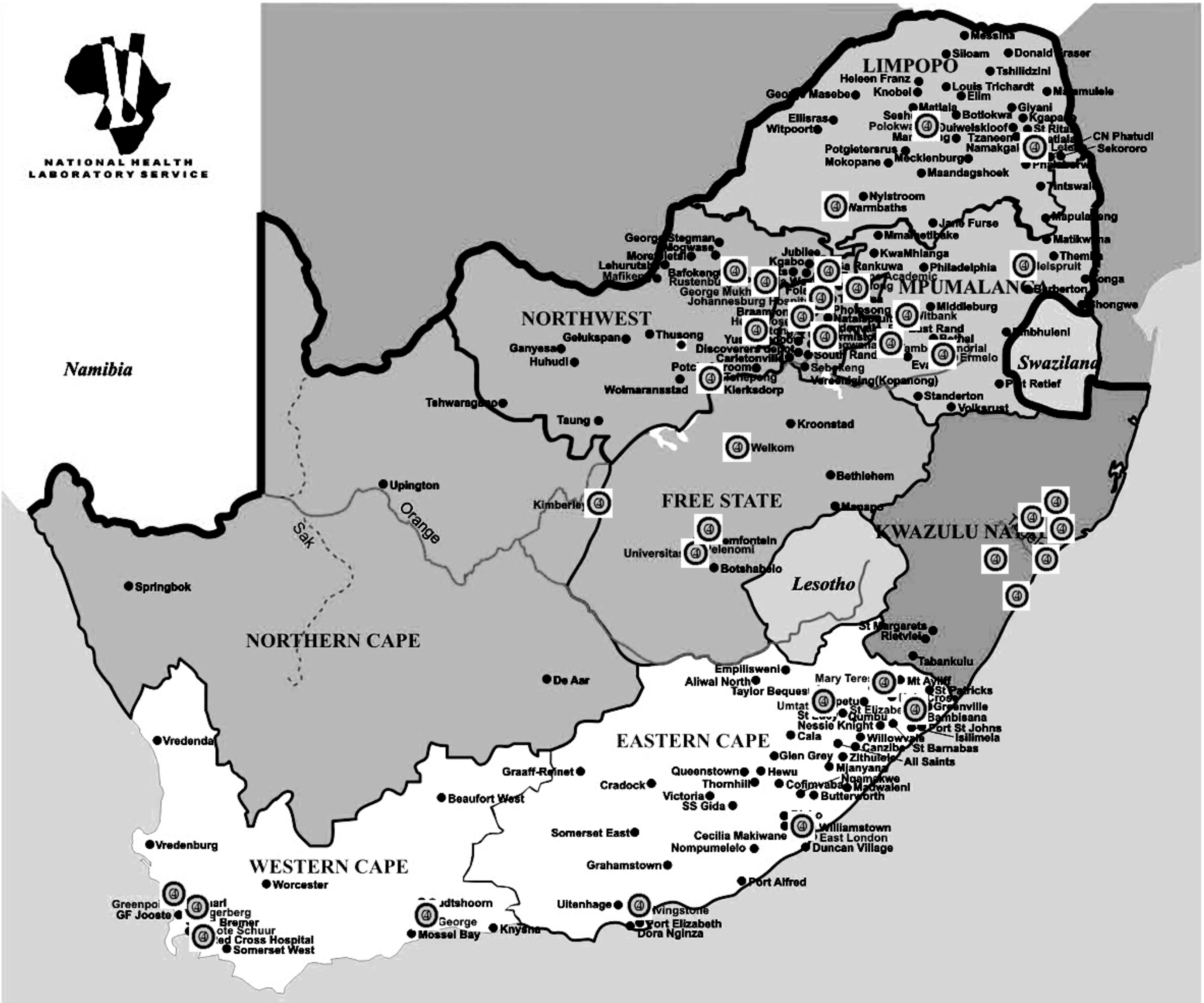
External validation for CD4 and viral load assays: South African Experience

W. Stevens

University of Witwatersrand and
National Health Laboratory Service


NATIONAL HEALTH
LABORATORY SERVICE
Namibia
LIMPOPO
NORTHWEST
MPUMALANGA
Swaziland
FREE STATE
KWAZULU NATAL
Lesotho
NORTHERN CAPE
EASTERN CAPE
WESTERN CAPE
Orange
Sak
Messina
Siloam
Donald Fraser
Tshilidzini
Louis Trichardt
Elim
Knobel
Helen Franz
Malamulele
Giyani
Botlokwa
Tzaneen
Namakgale
CN Phatudi
Sekororo
Ellisras
Witpoort
Potgietersrus
Mokopane
Mecklenburg
Maandagshoek
Tintswalo
Nylstroom
Warmbaths
Jane Furse
Mapulaneng
Matikwana
Philadelphia
Middleburg
Witbank
Embhuleni
Barberton
Shongwe
Piet Retief
Standerton
Volksrust
Thusong
Gelukspan
Ganyesa
Huhudi
Wolmaransstad
Klerksdorp
Taung
Tshwaragano
Kroonstad
Welkom
Bethlehem
Manapo
Upington
Kimberley
Universitas
Botshabelo
Springbok
De Aar
Vredendal
Vredenburg
Worcester
Somerset West
George
Mossel Bay
Knysna
Beaufort West
Graaff-Reinet
Cradock
Queenstown
Thornhill
Victoria
SS Gida
Somerset East
Cecilia Makiwane
Nompumelelo
Grahamstown
Port Alfred
Uitenhage
Port Elizabeth
Dora Nginza
East London
Duncan Village
Williamstown
Butterworth
Cofimvaba
Hewu
Glen Grey
Cala
Nessie Knight
Willowvale
Canzibe
St Barnabas
All Saints
Isilimela
Port St Johns
Bambisana
St Patricks
Greenville
Mary Teresa
Aliwal North
Empilisweni
Tabankulu
St Margarets
Rietvlei
GF Jooste

Methodologies evaluated in South Africa (NHLS/WITS)

CD4

- Abbreviated CD3/4/8 vs 6-tube NCCLS CDC panel *(Glencross et al)*
- Manual BC Cytospheres *(Glencross et al)*
- CD4 only/isotypic vs. vs 6-tube NCCLS/ CDC panel *(Sherman et al)*
- Imagn 9000 *(Glencross et al)*
- FACSCount, 4-color Trucount/Multiset, Ortho
  - Vs. BC Tetra *(Glencross et al)*
- DP PLG Vs. DP BC Tetra (Glencross et al, Forman et al)
- FACSCount, 4-color Trucount/Multiset, Ortho
  - DP PLG *(Glencross/ Scott et al)*
- SP PLG Vs. SP BC Tetra *(Glencross et al)*
- FACSCount, 4-color Trucount/Multiset, Ortho
  - SP PLG *(Glencross/ Scotts et al)*
- DP PLG Vs. SP PLG *(Glencross et al)*
- Dynabeads vs. PLG(SP/DP) & BC Tetra *(Tiounine/ Glencross et al)*
- GUAVA vs. PLG(SP/DP) & BC Tetra *(Scott et al)*
- PointCARE/ AuRICA (and versions) vs. PLG(SP/DP) & BC Tetra *(Scott/ Glencross et al)*
- Roche - KXpert vs. PLG(SP/DP) & BC Tetra *(Glencross/ Aggett et al)*
- Roche – pocHe vs. PLG(SP/DP) & BC Tetra *(Glencross/ Aggett et al)*

VIRAL LOAD

- HIV DNA Qualitatives *(Stevens, Sherman et al)*
- DBS - Discrete data *(Stevens, Sherman et al)*
- RNA Cobas AmpliPrep, Amplicor Monitor v 5 *(Stevens et al)*
- P24 Perkin Elmer *(Stevens et al)*
- RT (Cavidi) *(Stevens et al)*
- FRET (In house HIV-1 *(Rekhviasvili et al)*
- LUX (In house) *(Rekhviasvili et al)*
- NucliSens, EASY Q, miniMag and EasyMag *(Stevens, Scott et al)*
- ABBOTTM200RT *(Scott, Stevens et al)*

When to do Method Validation Studies?

- When considering purchasing a new system
- When placing a new system (equipment or kit) into service
- At regular intervals to assess on-going system performance
- When troubleshooting questionable system performance

  – **Balanced against cost!!**

Regulatory Guidelines for Method Validation:

- No regulatory authority in South Africa
- NHLS SOP
- Audited by SANAS
- **Reference documentation**:
- CLIA 88 Final Rule - (January 24, 2003)
- CAP General Checklist (July 2003)
- CAP Point-of-Care Checklist (July 2003)
- JCAHO - WT and QC Section (2004)
- NCCLS/CLSI – EP9A2,EP5A2,EP17A

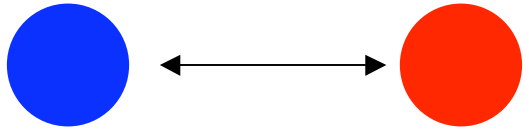
Clinical Laboratory Implementation

1. EVALUATION

Same reportable units with different methods, same site

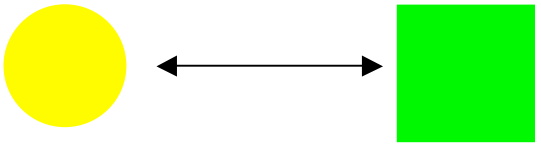
e.g. FACSCount vs. PLG CD4

both measuring CD4 cells/ul



Different reportable units with different methods, same site

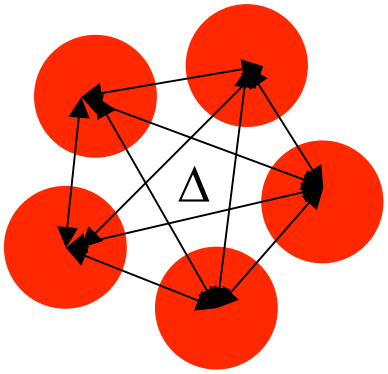
- e.g. p24 pg/ml vs. RNA copy number/ml



2. VALIDATION

Same method, different sites

- Local vs. International
- One-on-one
- Multiple sites: pool mean/median



GCLP (GCP+GLP) = 4 phases

- Phase I (background and set up)
  - Background and guidelines for scientific understanding of new method and determine characteristic of ‘outliers’
  - Define the ranges (units!!!) and known variability
  - Select appropriate technology where possible use:
    - Gold standard
    - More than one assay
    - More than one site
    - Use automation
  - Define the roles and responsibilities, plan ahead
  - Familiarity with instrument prior to implementation
  - Feasibility study

Protocol for Qualitative Assessment

- Volume of samples
- Published performance parameters
- Equipment needs
- Space requirements
- Workflow analysis
- Staffing requirements and skills
- Turnaround time
- Available IQC and EQA
- Cost per reportable
- Supplier availability
- Transport requirements
- Geographic locations of clinics

Study Readiness

- Professional Installation
- Appropriate staff training and device familiarization
- Ensure SOPs in place
- Consideration of sample selection, sample matrix, sample preparation
- Sufficient reagents within a lot
- Protocol familiarisation
- Records of continual PQ must exist
  - Calibration records
    - Frequency
  - Daily checking
  - Maintenance, cleaning
  - SOPs
  - Training
  - Data recording incl downtime, rejections etc

Components to an evaluation/validation protocol

- Precision
- Accuracy
- QC controls and reference material
- Reportable range and linearity
- Sensitivity
- Clinical interpretation
- Multi-site comparison

- Clinical longitudinal follow-up
- Switch study and change control
- EQA
- Others: selectivityreproducibility

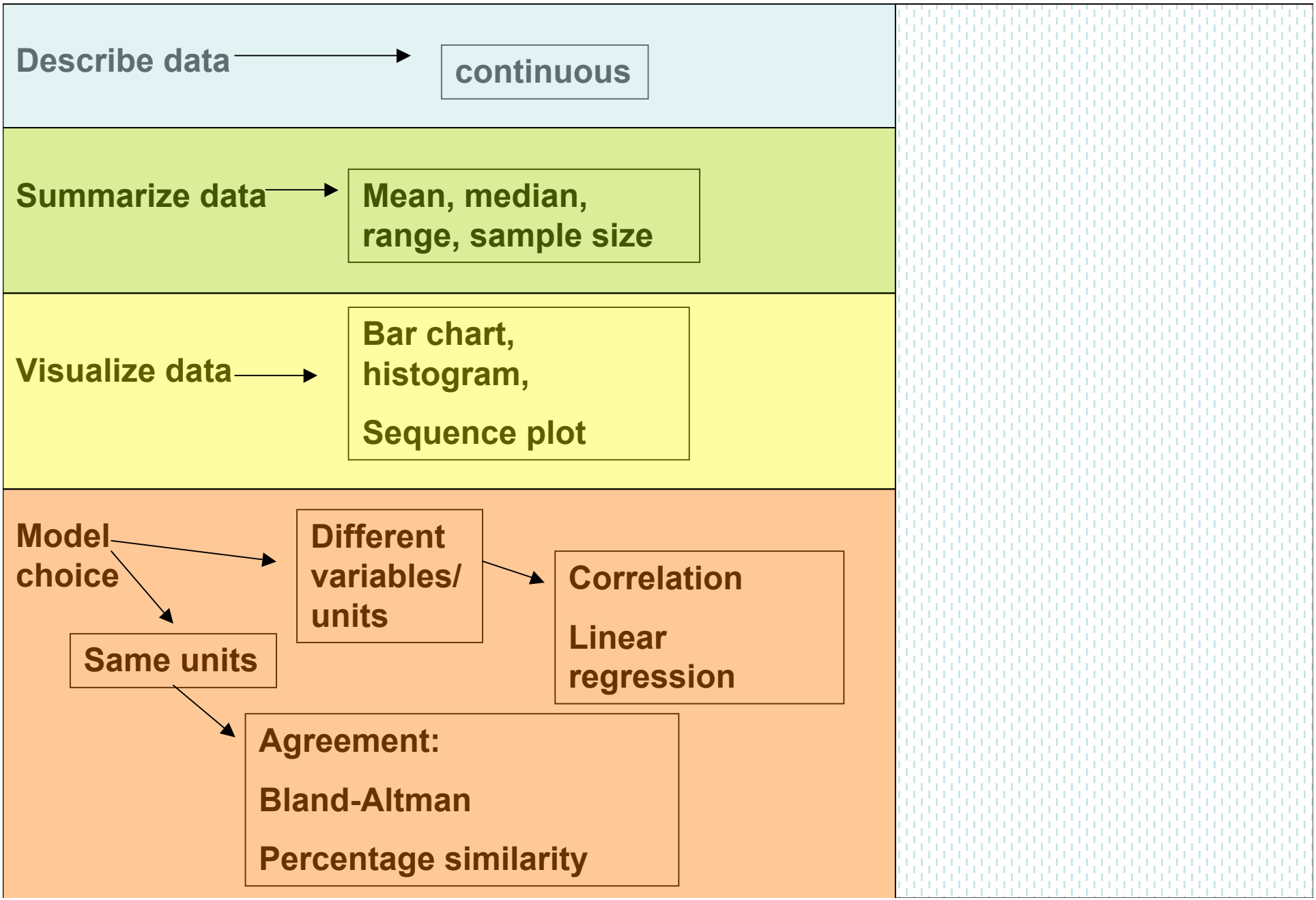
Phase II: Design and Analysis: Components for quantitative parameters

COMPONENTS	GUIDELINES	n	Viral load	allowable difference (with ref) and CV component	CD4	Allowable difference
Precision (instrument component)	CLSI EP5-A2, EP15-A2, EP21-A					
Within-run (same day/repeatability)	Strauss, 1996	10 Low, medium, high range	#	SD based on maufaturer's claims	#	<1%-6% CV
Between-run (several days/reproducibility)	Strauss, 1996	20 x 20 days (!)	#	SD based on maufaturer's claims	#	2.2%-4.3% CV
Accuracy	CLSI EP9-A2, EP21-A					
Patient samples over ranges		~40-100 (include all ranges)	#		#	
Bias determination (systematic error)	Strauss, 1996, Sherman1999		#		#	20cells/ul (Glencross), 25cells/ul (FACSCount), 30cells/ul (primary CD4)
SD of bias (random error)			#		#	
Overall (total analytical error) Variability for clinical significance	Martin 2000		#	0.3Log (intra), 0.5Log (inter), 1.0Log clinical		
QC controls and Reference material						
Within limits and across runs			#	VQA, BBI, WHO	#	NEQAS,QASI
Reportable (reference) range and linearity	CLSI EP6-A	1 sample/7-11 ranges (also use dilution)		~10% of value per range	#	200 - 350cells/ul
Sensitivity	CLSI EP17-A					
LoB (limit of blank)		20 HIV-	#	95% not detected	n/a	
LoD (Limit of Detection)		20 HIV+ at lowest range	#	95% must exceed LoD	n/a	
Clinical interpretation						
Impact on patient management	Martin 2000	all above	#	>1.0Log change is clinically significant	#	200 - 350cells/ul
Significance over range			#		#	
Multisite comparison						
Local			#		#	
International			#		#	
Clinical longitudinal follow up						
Impact on treatment over time		~20patients to week 12 (few reported studies)	#		#	
Switch study and change control						
EQA	VQA, NEQAS, QUASI	Annual long-term participation	#		#	

Acceptable Clinical Differences

- **CD4:**
  - ~20cells/ul @ 200cells/ul (NB: data range)
- **Viral load**:
  - 0.3 log copies/ml for intra-variability
  - 0.5 log copies/ml for inter-variability
  - 1.0 log copies/ml = clinical difference/patient mismanagement

Statistical approach: Method comparison protocol



Approaches to statistical analysis

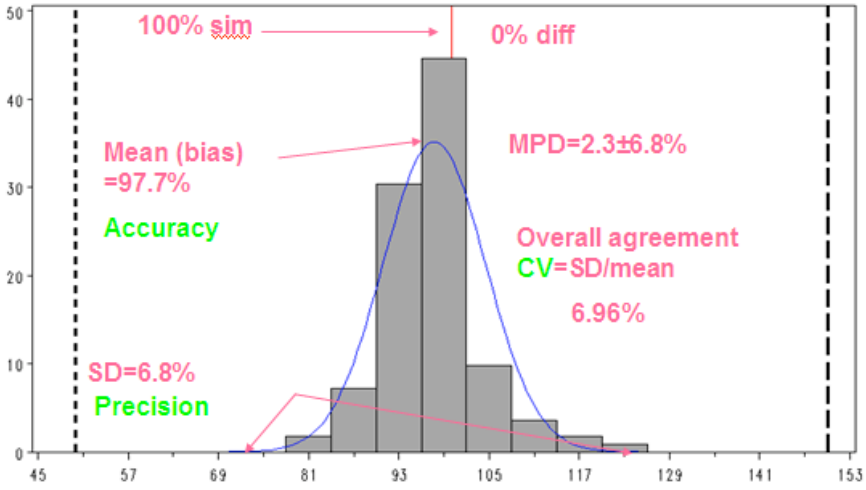
COMPONENTS	Statistical calculations and models applied	Visualization
Precision (instrument component)		
Within-run (same day/repeatability)	SD comparable to within manufacturers claims	Tables
Between-run (several days/reproducibility)	SD comparable to within manufacturers claims	
Accuracy		
Patient samples over ranges	Agreement	Scatter plots, histograms
Bias determination (systematic error)	Difference (Bland-Altman, 1986), Percentage Difference (Pollock,1992), Percentage similarity (Scott, 2003)	
SD of bias (random error)	Incl confidence intervals	
Overall (total analytical error) Variability for clinical significance	Percentage similarity CV (single unit of accuracy and precision)	Sequence scatter plots
QC controls and Reference material		
Within limits and across runs	SD and CV	
Reportable (reference) range and linearity	Linear regression, R2 coefficient of determination (emphasis on viral load)	
Sensitivity		
LoB (limit of blank)	Percentage pass	
LoD (Limit of Detection)		
Clinical interpretation		
Impact on patient management	Log transformations for viral load	
Significance over range		
Multisite comparison		
Local	Median and means	
International		
Clinical longitudinal follow up		
Impact on treatment over time		Box plots and line charts
Switch study and change control		
EQA		

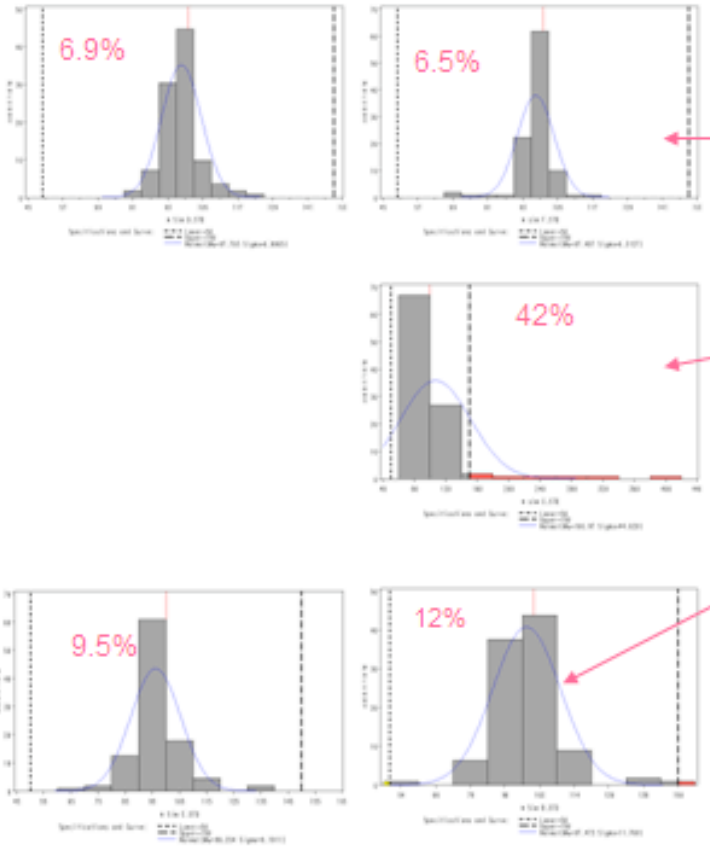
Percentage similarity (Scott et al, 2003)

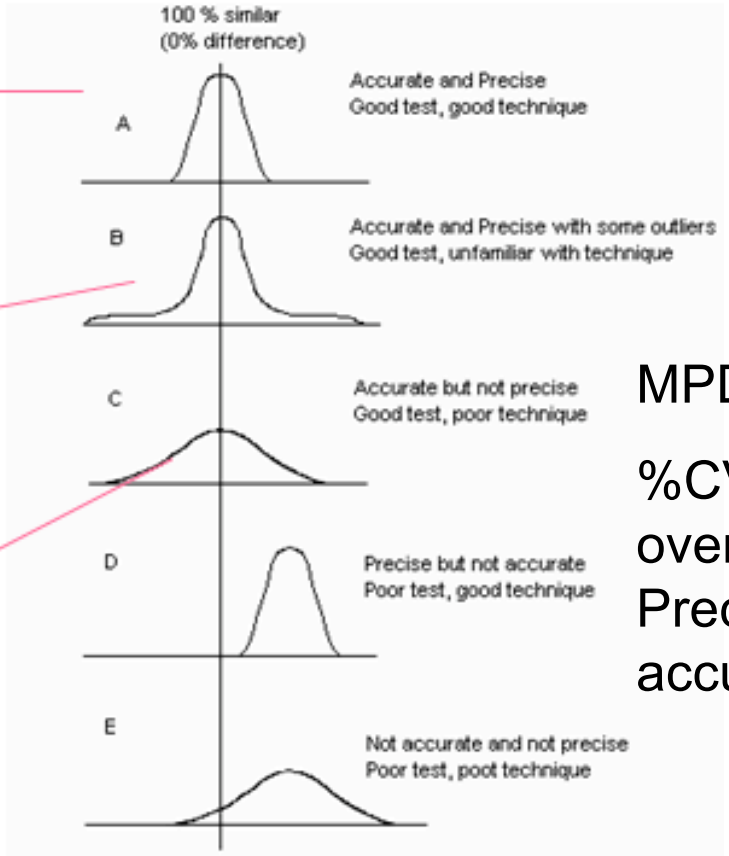
(((a+b)/2)/a) * 100, Where:

a=reference (predicate method)

B=new method







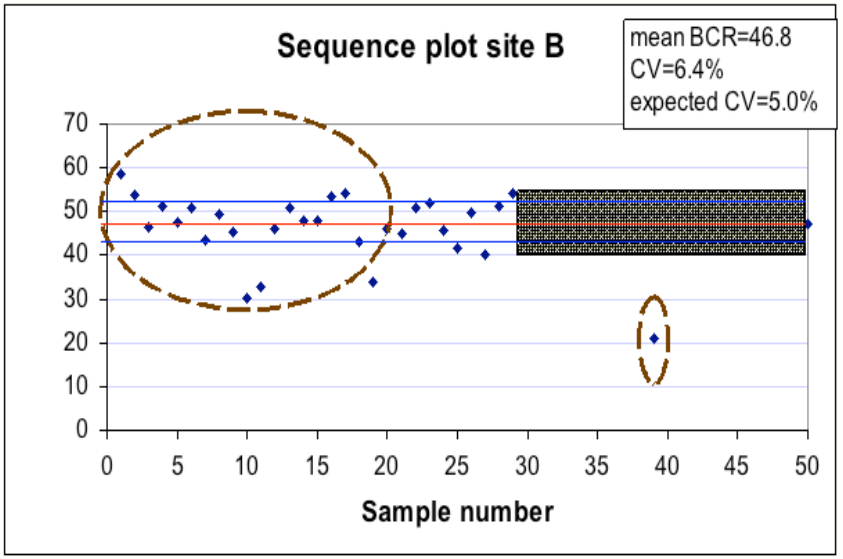
MPD=total error

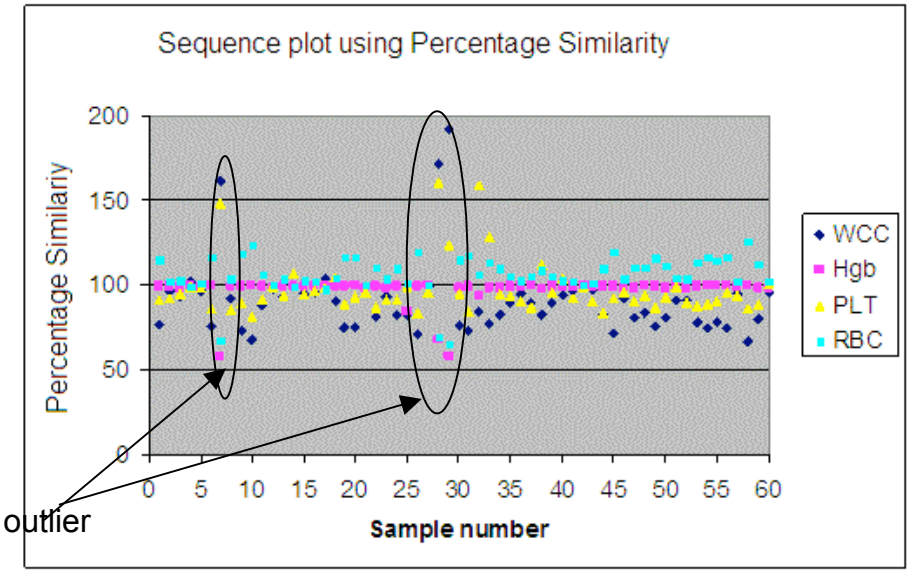
%CV = single unit of overall agreement of Precision and accuracy

Sequence analysis (Scott, 2005): measurement over time

- Similar to Levey-Jennings
- Scott et al (2005) developed using BCR for training and measuring pipetting skill.

- Mostly analyse data in relation to the range of values.
- Now plot the difference or percentage similarity (advantage for multiple parameters and multiple methods) on the vertical axis in sequence of sample preparation – horizontal axis
- Random scatter vs trends, clusters can explain outliers and instrument performance.





Sample integrity not real outlier

Average vs absolute for determining the bias over the range

- Bland-Altman (1995) prefer using average to reduce artifacts and misleading differences.
- This may be advised for small ranges in data, but our experience on broad data ranges such as CD4 is to suggest using the absolute reference method
- Since the values measured come from reportable results.
- A new method is never reported in the clinical context as an average (handled by change control and switch studies)
- Examples: clinical reportable results for patient management
  - Viral load:
    - A=400c/ml, B=16000c/ml, average = 8200c/ml is above clinical range
  - CD4:
    - A=150cells/ul, B=600cells/ul, average = 375cells/ul. Above clinical range 200cells/ul.

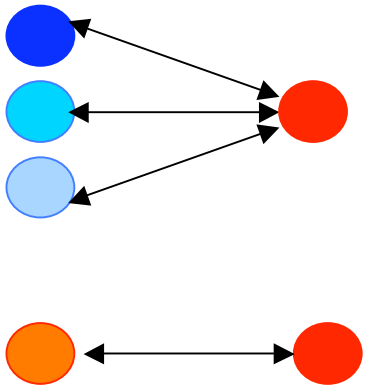
Phase III and IV

- Phase III: reporting
  - **Validation report** including robustness, ease of use, downtime
  - Good documentation
  - Store everything!!!
  - Take action if deviations
- Phase IV: Follow-up
  - Accepted into routine and do follow-up
  - Long-term stability
  - Gather in-study data
  - Handle change control
  - Participate in EQA programmes/proficiency testing

Clinical Laboratory Implementation
PLG CD4 case study

1. Method EVALUATION
   - DP PLG vs. International guideline methods

     including an international site
     - (Glencross, Scott, Jani, Barnett Janossy, Cytometry, 2002)
       - International - Ortho Cytoron (volumetry) with TRANSFIX™
       - Local - BD 4-colour Trucount™ - SP
       - Local - BCI 4-colour TetraONE™ - SP and DP



   - Cross platform evaluation

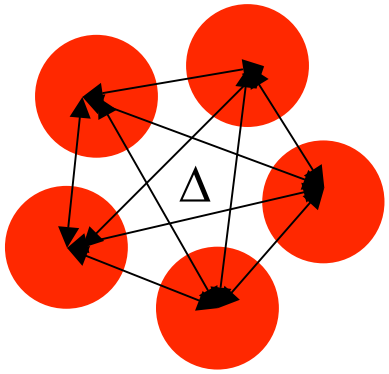
     SP vs. DP - PLG CD4

2. International/ Independent evaluation

   including multi-site testing

   DP PLG and SP PG

   1. Storie et al (UK),
   2. Pattanapanyasat et al (Thailand)
   3. Chianese et al (Italy)
   4. Denny et al (US)
   5. Mandy et al (Uganda)
   6. Forman et al (US/ China)
   7. Walker et al (UK)
   8. Ceffa et al (Mozambique)
   9. Romano et al (Italy)



3. Long-term CLINICAL VALIDATION: Longitudinal follow-up-CIPRA cohort

Comparing international predicate type methods with PLG both as SP and DP
(2 years follow-up to date >>> 5 years).
Follow-up on 711 patients across 2 geographically separate sites with >90% followed to 2 years.

Glencross et al, in press

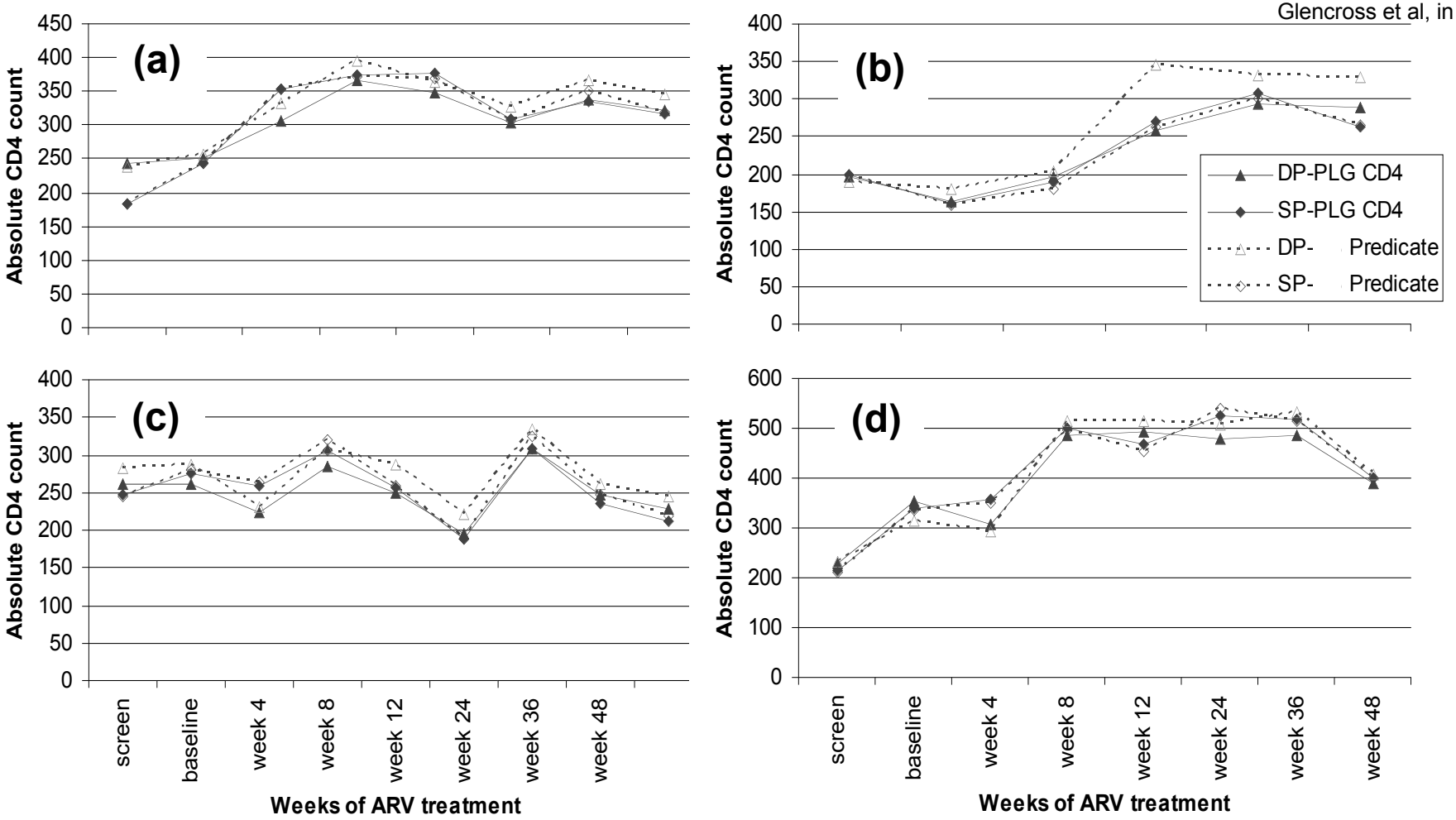
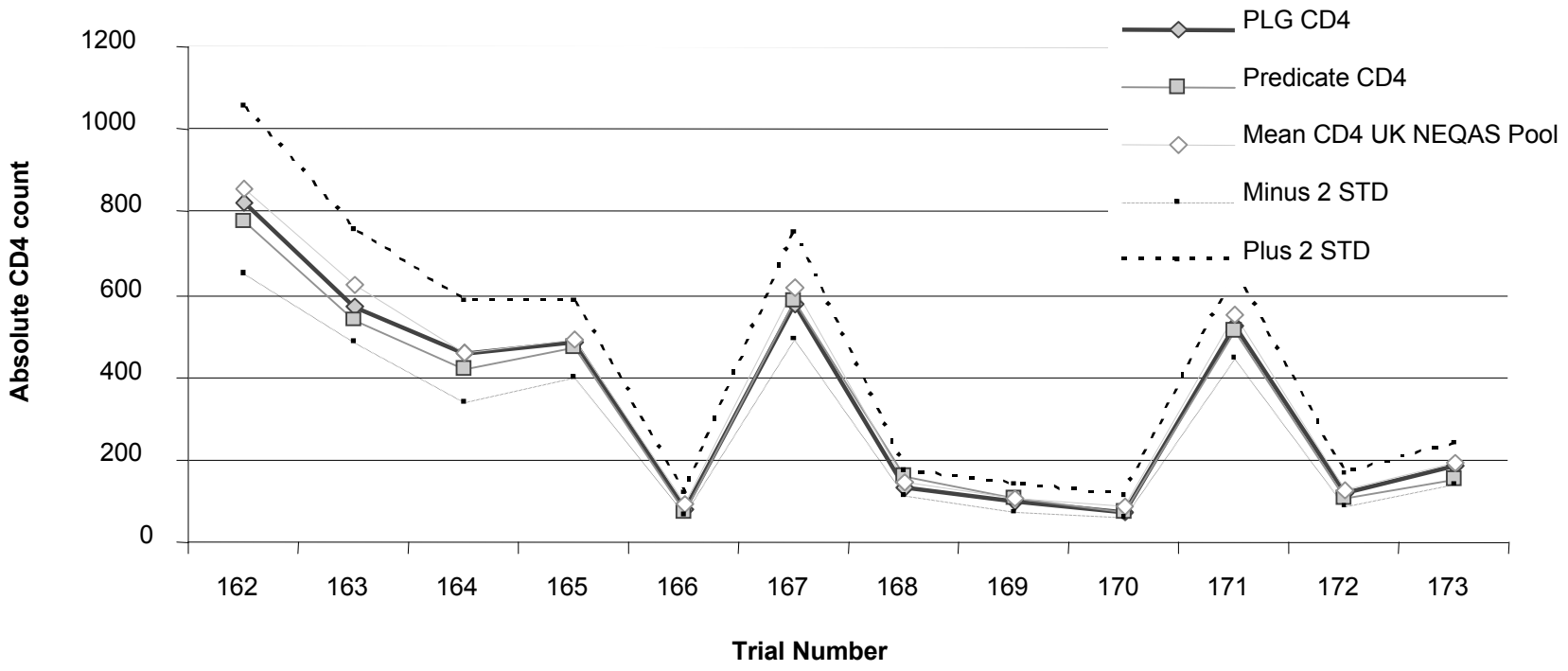


FIGURE: Four random patient examples showing longitudinal follow-up across US Predicate and PLG as both SP and DP

4. International EQA validation: ACCURACY

performance of PLG versus US Predicate with UK NEQAS



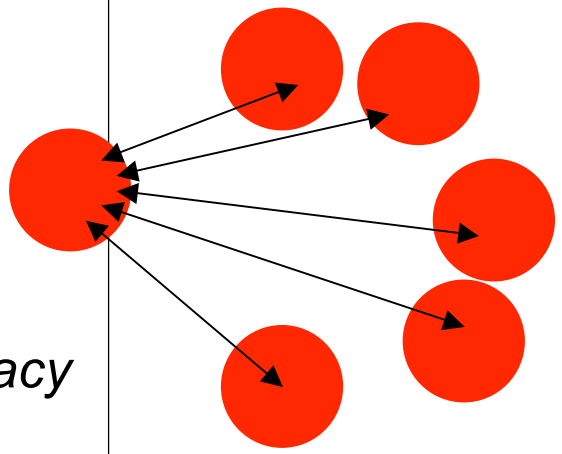
Glencross et al, in press

5. National NHLS laboratory implementation

Site by site validation with NHLS reference site

BA, %SIM analysis
to assess accuracy
Within-laboratory precision <5%
SP and FCR ensure pipetting accuracy
Workshops and on-site training

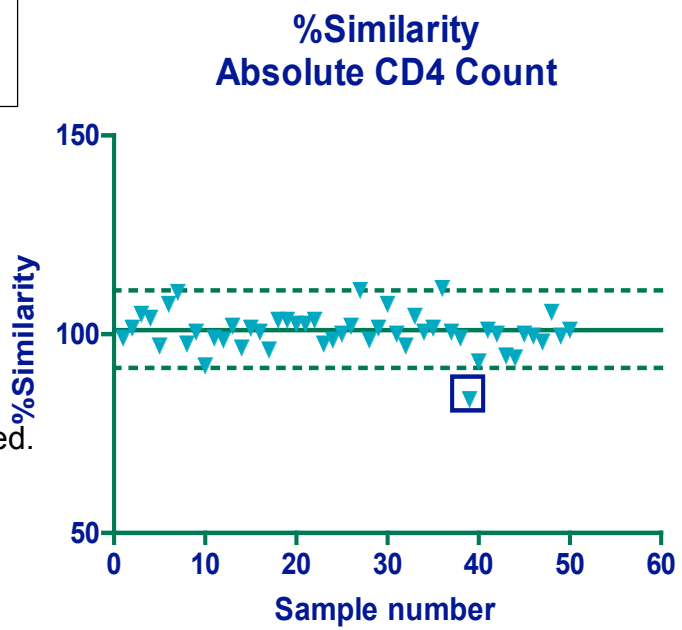
Initially 22 labs validated
35 participating labs – June 2006
52 participating labs – June 2007



Sample	XL1	XL2	%Similarity	Clinical Significance
39	2653	1791	83.7	NCS

SITE A comparison to Reference LABORATORY: One possible outlier was identified.

Removing this sample from the analyses did not significantly change the % similarity (Mean of 100.9±4.16 with a CV of 4.13%), indicating the excellent overall agreement for this parameter between instruments.



6. Ongoing external CLINICAL VALIDATION: 2-monthly EQA

Performance of 52 NHLS participating PLG CD4 sites, in support of the National ART programme

EQAS validation by participation in the NHLS CD4-AFREQAS (>250 labs)

(2/6 trials organized in collaboration with *WHO* and *QASI*)

Data presented at AIDS XVI Toronto (in press)

Performance of NHLS PLG CD4 laboratories as an organized national network versus the performance of laboratories participating in other African laboratories or internationally on the QASI programme

Between-lab precision - %CV

Fig.8 Glencross et al.

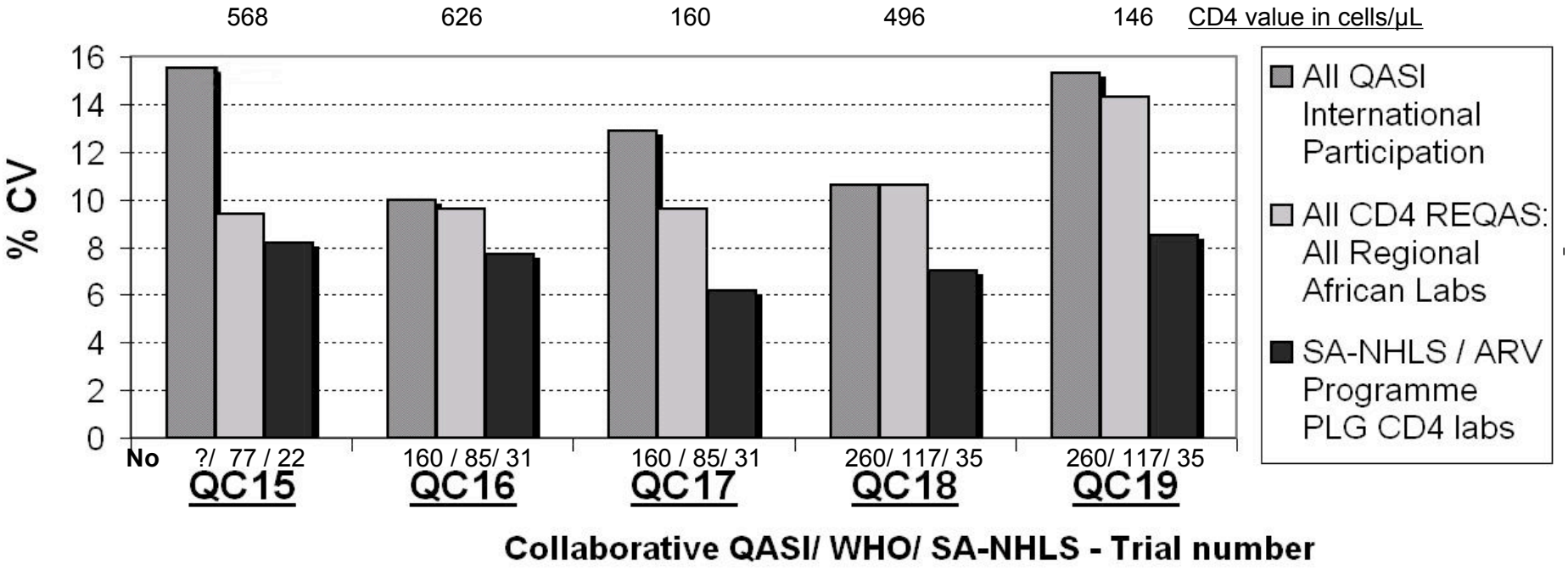
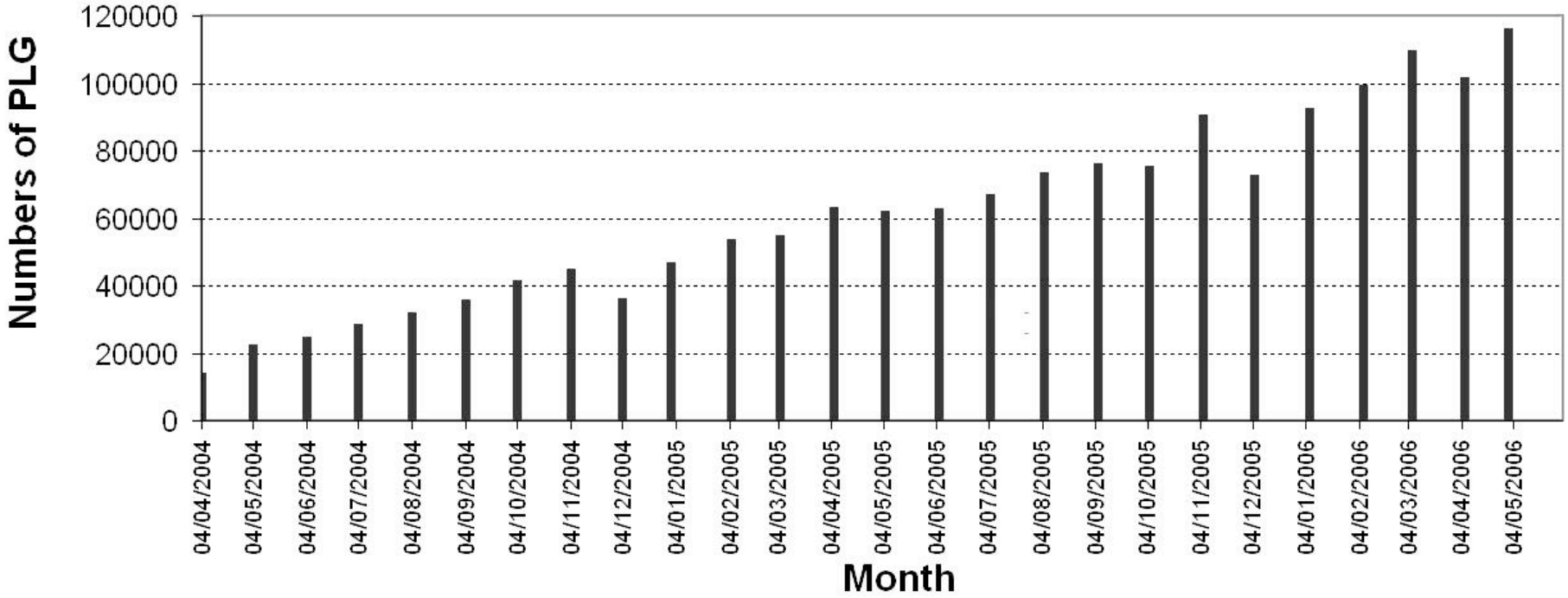


Fig.10 Glencross et al.



Glencross et al, in press

Acknowledgements

- Lesley Scott (University of the Witwatersrand, Johannesburg)
- Debbie Glencross (NHLS, SA)

- Web site launched: www.methodcomparison.com
- Training manual and workshop on method validation