Technical Brief on CD4 Technologies
(May 2010)

Contents

1.0 Aims and objectives
2.0 Target audience
3.0 Introduction
4.0 Background
  4.1 Physiological variability in CD4 count
  4.2 CD4 as a prognostic marker for HIV progression
  4.3 WHO recommendations for the use of CD4 monitoring in the management of adults and adolescents living with HIV
    4.3.1 The role of CD4 testing in deciding when to start ART
    4.3.2 The role of CD4 testing in deciding when to switch to second line ART
5.0 Description of available technologies for CD4 enumeration
6.0 Considerations when choosing CD4 technology
  6.1 Operating characteristics
  6.2 Performance characteristics
    6.2.1 Accuracy
      6.2.1.1 Measures of test accuracy
    6.2.2 Precision
7.0 Systematic review of performance of CD4 technologies
  7.1 Methods
    7.1.1 Inclusion criteria
    7.1.2 Data extraction
    7.1.3 Assessment of study quality
  7.2 Results
    7.2.1 Accuracy
      7.2.1.1 Bias/mean difference
      7.2.1.2 Misclassification probabilities
    7.2.2 Precision
      7.2.2.1 Within-laboratory precision
      7.2.2.2 Between-laboratory precision
    7.2.3 Conclusions of systematic review
8.0 Importance of EQA programmes and QC
9.0 Suggested study design for future evaluations of CD4 technology
10.0 What is on the horizon: new CD4 technologies
11.0 References

Appendix A: Search strategy and search terms
Appendix B: Algorithm for study inclusion
Glossary

Accuracy

A general term used to describe how similar the results obtained for individual specimens with a new CD4 assay are to the results obtained with a reference CD4 assay. Measures of accuracy include bias and misclassification probabilities.

Bias

A measure of how much the CD4 count that is obtained using a new assay differs from the CD4 count that is obtained using a reference assay when the same specimen is analysed using both assays. (bias need not be constant over the clinically important range of CD4 counts, and need not be the same for different laboratories using the same reference and new technologies.)

Coefficient of variation

The coefficient of variation (CV) is usually defined as the ratio of the standard deviation to the mean, which is usually estimated as the standard deviation divided by the average. This is often multiplied by 100 and expressed as a percentage (CV%).

Mean

The average value of a random variable, X. The most common estimate of the mean is the average of all the observations.

Median

The value m (of a continuous random variable, X) such that the probability that X is less than m is one half.

Misclassification probabilities

The upward and/or downward misclassification probability can be calculated for a particular cut-off, C. Among specimens with a CD4 count (obtained using the reference technology) which is less than C, the upward misclassification probability is defined as the percentage of specimens for which the CD4 count (obtained using the new technology) is greater than or equal to C (this is analogous to definition of 1 minus the sensitivity in a diagnostic test with only two values: the patient has the disease and the patient does not have the disease). Among specimens with a CD4 count (obtained using the reference technology) which is greater than or equal to C, the downward misclassification probability is defined as the percentage of specimens for which the CD4 count (obtained using the new technology) is less than C (this is analogous to definition of 1 minus the specificity in a diagnostic test with only two values: the patient has the disease and the patient does not have the disease).

Precision

Precision (also known as reproducibility) is a general term for how close the results on replicate specimens are (using a single assay technology). Several sorts of precision can be relevant, including within-technician, between-technician, within-laboratory and between-laboratory precision. Measures of precision include the coefficient of variation (CV%).
1.0 Aims and objectives

To provide updated technical guidance to optimize the use of CD4 technologies for HIV care.

To provide an evaluation of the performance characteristics (accuracy and reproducibility of results) of available CD4 technologies through a systematic review of published literature.

To review the operational characteristics of currently available CD4 technologies, and outline new technologies under development.

To provide guidelines for conducting performance evaluations of new CD4 technology, using clinically relevant measures of test performance.

2.0 Target audience

Ministries of health; national HIV/AIDS control programmes; laboratory managers and experts; manufacturers of CD4 technologies; procurement agencies

3.0 Introduction

HIV has particular tropism for CD4 cells and destruction of CD4+ cells results in many of the deleterious effects of HIV. Measurement of CD4+ T lymphocytes in peripheral blood is therefore a critical laboratory parameter for the evaluation and monitoring of patients with HIV. It is used to determine the immunological stage of HIV infection, for identifying when to start antiretroviral therapy (ART), to identify patients likely to benefit from co-trimoxazole prophylaxis and to recognise those most at risk of developing immune reconstitution syndrome. In addition, CD4 may be used in evaluating the response to treatment and for recognizing treatment failure.

WHO therefore encourages national programmes to increase access to CD4 measurement technologies to improve the quality of HIV prevention, care and treatment programmes.

A number of techniques for determining CD4 exist, however the range of cost, feasibility, ease of use, accuracy and reproducibility vary considerably. In general the cost and complexity mean that access to CD4 and viral load measurement is still highly restricted in many low income countries. The infrastructure of clinical laboratories located in many developing countries is often inappropriate to support routine determination of CD4 and viral load measurements, and high costs of equipment, reagents and problems associated with machine maintenance further restrict access.

It is unclear what proportion of adults and children living with HIV have access to laboratory monitoring tests such as CD4 and viral load given the cost, complexity and lack of capacity to perform monitoring tests in many resource limited settings. However significant investments have been made in developing laboratory capacity and countries are increasingly gaining access to these technologies.
4.0 Background

CD4+ T-lymphocytes, also known as the helper T-cells, are the coordinators of the immune response which protects the body against microbial disease and some forms of cancer. The destruction of CD4+ T-lymphocytes by HIV is the main cause of the progressive weakening of the immune system in HIV infection, and leads ultimately to the acquired immune deficiency syndrome, AIDS.

4.1 Physiological variability in CD4 count

Diurnal variation in CD4 has been observed in HIV-negative and HIV-positive subjects, with CD4 counts typically lower in the morning than the evening\textsuperscript{51,52}. Week-to-week and seasonal variation may occur, and changes in CD4 count may also be seen related to exercise, smoking, intercurrent infection, and some medications other than ART (eg corticosteroids). Changes in total white cell count and lymphocyte count may contribute more to the fluctuations in absolute CD4 count seen in response to these biological factors than does the CD4\%\textsuperscript{51,54}. These physiological intrasubject variations in CD4 count may account for as much as, if not more than, technical variability of CD4 measurement\textsuperscript{53}.

4.2 CD4 as a prognostic marker for HIV progression

CD4 count is a strong predictor of HIV progression to AIDS and death, and is by consensus considered the best laboratory marker for deciding when to initiate antiretroviral treatment (ART)\textsuperscript{54,55}. There is much heterogeneity in rate of CD4 decline, with a minimum period between seroconversion and AIDS of about 2 years, and a median of 11 years\textsuperscript{54}. In addition, the rate of decline in CD4 count for an individual patient may be highly variable over time\textsuperscript{56}, making it difficult to predict at higher CD4 counts the likely time course of an individual’s progression to advanced disease.

The use of clinical staging alone to determine ART initiation is limited by the uncertain prognostic value of asymptomatic or mild disease. In a Malawian setting, 56% of 1663 patients with WHO clinical stage 1 or 2 disease had CD4 counts of <350 cells/µl, and 36% had CD4 counts <250 cells/µl\textsuperscript{57}. The authors also reported high rates of attrition from healthcare among those patients not started on ART; it may therefore be supposed that, were CD4 testing not performed, many patients unrecognised as eligible for treatment may progress to advanced disease or die before re-presenting to healthcare.

Regular monitoring of CD4 counts, in conjunction with clinical criteria, is therefore recommended for monitoring HIV progression, with treatment initiation indicated when CD4 drops to below a defined threshold.

4.3 WHO recommendations for the use of CD4 monitoring in the management of adults and adolescents living with HIV\textsuperscript{58}

4.3.1 The role of CD4 testing in deciding when to start ART

In November 2009, the WHO issued a rapid advice update on their Antiretroviral therapy for HIV infection in adults and adolescents guideline\textsuperscript{58}. A clear stimulus to this update was the increasing evidence of improved survival and reduced HIV-related illnesses with earlier initiation of antiretroviral therapy (ART). A key recommendation of the updated guideline is therefore that ART
be started in all patients with HIV who have a CD4 count ≤350 cells/µl, irrespective of clinical symptoms. As a result of the uncertain prognostic value of WHO clinical stages 1 and 2, the guidelines recommend that all HIV-infected individuals with WHO clinical stage 1 or 2 disease should have access to CD4 testing to decide if treatment should be initiated.

For all patients with WHO clinical stage 3 or 4, ART initiation is recommended irrespective of CD4 count.

<table>
<thead>
<tr>
<th>WHO recommendations for when to start ART</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start antiretroviral treatment in all patients with HIV who have CD4 count &lt;350 cells/mm³ irrespective of clinical symptoms. (Strong recommendation, moderate quality of evidence)</strong></td>
</tr>
<tr>
<td><strong>CD4 testing is required to identify if patients with HIV and WHO clinical stage 1 or 2 disease need to start antiretroviral treatment. (Strong recommendation, low quality of evidence)</strong></td>
</tr>
<tr>
<td><strong>Start antiretroviral treatment in all patients with HIV and WHO clinical stage 3 or 4 irrespective of CD4 count. (Strong recommendation, low quality of evidence)</strong></td>
</tr>
</tbody>
</table>

The criteria for initiating ART for pregnant women are the same as for non-pregnant adults, with an antenatal CD4 count considered critically important to determine eligibility to ART for the health of the woman. Shorter term ART prophylaxis for prevention of mother to child transmission is indicated when the mother does not meet the criteria for treatment for her own health.

### 4.3.2 The role of CD4 testing in deciding when to switch to second line ART

Where available, it is recommended that viral load be used for confirmation of treatment failure. However when viral load is not available, the use of immunological criteria is recommended to confirm clinical failure, having been shown in 2 randomised controlled trials to be superior to clinical monitoring alone.⁵⁹,⁶⁰
5.0 Description of available technologies for CD4 enumeration

The gold standard technology for CD4+ T-cell counting is flow cytometry, a technology that allows a single cell to be measured for a variety of characteristics. Flow cytometers are used in cell counting or cell sorting applications.\textsuperscript{44} For CD4+ T cell enumeration, cells are stained (fluorescence tagged cell surface markers) for CD4+ T-cell counting. The standard methodology for determining the percentage of CD4+ cells is by staining for CD45/CD3/CD4. By first selecting the lymphocytes (CD45), the CD3/CD4 cytogram can be used to enumerate CD4+ cells. More extensive or simplified staining panels are commercially available.

While flow cytometry generally gives the percentage of CD4+ or CD8+ cells, some flow cytometers precisely record the volume of sample analysed or deliver a fixed volume of sample. This method is called volumetric approach and can directly convert CD4% to absolute CD4 counts. In instruments without this facility, a fixed volume of sample is spiked with a known number of fluorescent beads (beads-based systems). Since the fluorescent beads can easily be differentiated from cells by the flow cytometer, the counted beads of the sample analysed permits volume determination and hence the concentration of cells. The volumetric approach or systems using fluorescent beads are called \textbf{single platform} approaches and are commercially available as TruCount (Becton Dickinson, USA), FlowCount beads (Beckton Coulter) and Perfect Count (Cytognos) that can be used on practically any flow cytometer. Besides using the traditional flow cytometers (open platforms), dedicated (often called closed) platforms are developed for CD4+ T-cell counts. The three commercially available dedicated single platforms are: FACScount (Becton Dickinson), Guava easy CD4 (Guava technologies) and Partec Cyflow Counter (Partec). The dedicated platforms in comparison to flow cytometers have decreased applicability, but allow CD4+ T-cell counting with reduced technical complexity.

Another approach used to calculate absolute CD4+ counts is to use a haematology analyser. The haematology analyser measures the total white cell count. This \textbf{dual platform} approach introduces variability in absolute CD4+ counts due to combining results of two platforms into a single calculation. Recent developments in staining panels and data analysis (Panleucogating) offer better results over traditional dual platform methodology, increasing the performance of the dual platform approach up to the single platform level\textsuperscript{45}.

Besides the traditional flow cytometry methodology, manual methods exist. Two manual methods use microscopy (fluorescence or light) and haemocytometers. By using beads (Dynal T4 Quant, Dynabeads) or spheres (Coulter Manual CD4 count kit, Cytospheres) that become distinctly visible by microscopy when in contact with CD4+ T-cells, it is possible to count cells.
6.0 Considerations when choosing CD4 technology

In deciding upon a technology, both the operating characteristics and the performance characteristics will require careful consideration.

6.1 Operating characteristics

Operating characteristics include:

- Environmental requirements of the technology: instrumentation, power supply, refrigeration of reagents, robustness to heat and humidity
- Parameters measured – absolute CD4, CD4%
- Ease of use, training required, degree of automation
- Cost
- Throughput
- Compatibility with external quality assessment (EQA) programmes and availability of quality control (QC) reagents
- Compatibility with stabilised specimens

The operating characteristics of the currently available CD4 technologies are presented in Tables 1-4.

Desirable operating characteristics will differ according to the level of the healthcare system, and are likely to be very different at primary level health posts than at national reference laboratories. A tiered approach reflecting the capabilities of, and demands on, each level of the health care system could assist in the allocation of limited resources. The desirable operating characteristics for CD4 testing are described for each level of the health care system in Table 5.

Other considerations in the context of rapid expansion of testing capacity include the existing instrument base and laboratory infrastructure, the expected changes in testing demand over the subsequent 2-3 years, the availability of human resources, and the need for CD4% for paediatric testing.

Compatibility with EQA programmes and availability of QC reagents are of prime importance and are discussed further in section 8.0.
6.2 Performance characteristics

In addition to considering operating characteristics when choosing a technology, it is essential to consider performance characteristics. Performance in this context refers to the capability of a technology to give accurate and reproducible results, and therefore to guide patient management appropriately.

In evaluating the performance of a test, both accuracy and precision must be considered.

6.2.1 Accuracy

Accuracy is a measure of the degree of closeness of the reported value to the true value. Accuracy of a new technology is evaluated by comparing results generated by the new test with those obtained for the same samples using a reference technology.

6.2.1.1 Measures of test accuracy

Correlation and linear regression analyses are frequently used to report accuracy, but this approach is inappropriate in the evaluation of CD4 technology. In this context, correlation coefficients do not indicate level of agreement, and may be misleading. Unacceptably high levels of misclassification (a clinically relevant measure of accuracy) may be seen despite excellent correlation. For these reasons, assessing correlation alone is insufficient.

Other measures of accuracy include bias and misclassification.

Bias or mean difference reflects the average difference between the results of the technology under evaluation and an accepted reference standard. Bias may be reported using Bland-Altman analysis, in which the difference between two methods of measurement is plotted against their mean. If the differences do not vary systematically over the range of CD4 counts studied, then the overall mean difference can be calculated and indicates bias. If the differences are normally distributed, then 95% of the differences will lie within 2 standard deviations (referred to as ‘limits of agreement’). If these limits are clinically acceptable, agreement between the two methods is adequate, and the new technology may be used in place of the reference standard. However, in CD4 measurement, the magnitude and variability of the differences between two methods are nearly always larger at higher counts, making this approach unreliable. Even within the clinically important range of CD4, bias may not be constant.

Misclassification probabilities describe the likelihood that a test will incorrectly categorise a result as higher or lower than a given cut-off value. They have the advantage of being a clinically relevant measure of accuracy, as they can be used to assess the likelihood that a patient will be incorrectly classified above or below a defined CD4 threshold used in clinical decision making.

Two types of misclassification can be defined — upward misclassification probability and downward misclassification probability. Among specimens with a CD4 count (obtained using the reference technology) which is less than a defined cut-off of value, upward misclassification probability is defined as the percentage of specimens for which the CD4 count (obtained using the new technology) is greater than or equal to that cut-off. Among specimens with a CD4 count (obtained using the reference technology) which is greater than or equal to the cut-off, the downward misclassification probability is defined as the percentage of specimens for which the CD4 count (obtained using the new technology) is less than the cut-off.
Misclassification probabilities are therefore analogous to $1$ minus the sensitivity and $1$ minus the specificity of a test around a determined value.

Upward misclassification around a treatment threshold may be most clinically important (leading to delay of start of ART or prophylactic treatment in some patients). Downward misclassification may result in treating large numbers of patients who have CD4 counts above the guideline threshold.

A disadvantage of misclassification probabilities as defined here is that they do not differentiate between low-magnitude inaccuracy, e.g. a count of 349 cells/µl being misclassified as >350 cells/µl, which may be reasonably expected from any test, and high-magnitude inaccuracy. Misclassification probabilities are therefore heavily influenced by the study population. A study which includes many CD4 values falling close to the defined cut-off (as measured by the reference test) will show higher rates of misclassification by the new test than a study where the majority of values lie away from the threshold. For this reason, reporting the range and distribution of CD4 values of the study population is of great importance.

A more important measure might be the probability that a patient with a CD4 test well below the treatment threshold might be incorrectly classified as above that threshold, e.g. the probability that a patient with a ‘true’ CD4 value of <250 cells/µl might be misclassified as >350 cells/µl. Unfortunately these data are currently rarely available in the published literature.

### 6.2.2 Precision

Precision measurements answer the question ‘how close are the results of replicate testing using a single technology?’ Precision is particularly important when following a patient’s serial measurements using the same technology. Precision can be measured within-run, between-run, or between-laboratory.

Within-run variability is calculated from testing multiple replicates of the same specimen within a single machine run. Between-run variability is calculated from testing multiple replicates of the same specimen in different runs or at different times. The two measures together describe the amount of variation that can be expected among results using the same specimen. Precision is also called reproducibility and can be thought of as the degree to which repeated measurements under unchanged conditions show the same results. Precision can also be calculated between laboratories to determine how much variation is introduced into the results by having the test performed by different technicians in different settings (between-operator or between-laboratory variability).

Data on precision are often reported as the coefficient of variation (CV) which is a measure of dispersion similar to standard deviation but is calculated with consideration for the data and can therefore be compared across studies. A lower value of CV indicates less variation and greater assay reproducibility. %CV may be badly underestimated if based on too few replicates; a minimum of 8 replicates should be used.

Manufacturers publish information on their technology’s accuracy and precision; however, independent, peer-reviewed evaluations are a more reliable source of information. Evaluations of technology should include a comparison of the test under evaluation to a reference standard of acceptable standing in the scientific community. Before deciding to implement a technology, sites
need to be satisfied that independent peer-reviewed evaluation data have shown acceptable performance.
7.0 Systematic review of performance of CD4 technologies

We performed a systematic review of studies evaluating the performance of CD4 enumeration technologies. A search of the Cochrane Library and the Centre for Reviews and Dissemination databases (including DARE, NHS EDD and HTA) found no existing reviews addressing the review objective.

7.1 Methods

We followed standard guidance in performing the review47. Details of search strategy, search terms and algorithm for study selection are included in appendices A and B.

7.1.1 Inclusion criteria

Inclusion criteria are shown below.

<table>
<thead>
<tr>
<th>INCLUSION CRITERIA (PICOS)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td>Include studies of performance of CD4 enumeration technologies in HIV infected and/or non-infected adults, in any geographical location</td>
</tr>
<tr>
<td><strong>Interventions (index tests)</strong></td>
<td>Include any commercially available technology used for the enumeration of CD4 count, including (but not limited to) manual bead-based technologies, flow cytometry systems requiring minimal technical input, and flow cytometry systems requiring significant operator input or flow cytometry skill</td>
</tr>
<tr>
<td><strong>Reference standard/comparators</strong></td>
<td>Include studies using any acceptable (currently or at the time of the study) flow cytometric method as the reference standard</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Include studies evaluating accuracy (including e.g. bias and/or misclassification probabilities) and/or precision. Include studies calculating sensitivity, specificity, PPV and NPV for a defined CD4 cut-off value</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Include evaluation studies using an acceptable reference technology</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Include English language only, 1990-present</td>
</tr>
</tbody>
</table>

Currently, no gold standard technology or internationally recognised reference preparation exists for CD4, and a wide range of flow cytometric methodologies have been used as comparators. For the purposes of this review, we included studies using as reference technologies any flow cytometric methodology considered by the review panel to be acceptable methodology currently or at the time of the study.
7.1.2 Data extraction

The following details were extracted from studies accepted for inclusion: location of study, laboratory or health care level, degree of manufacturer involvement in the study, index test, reference test, population (HIV positive or HIV positive and negative). The following measures of accuracy were extracted: bias (mean or median difference), including limits of agreement if Bland-Altman analyses were performed, and maximum observed difference when available; misclassification probabilities around a defined CD4 cut-off (when sensitivity or specificity values were given, misclassification probabilities were calculated). If data were available for subgroups of CD4 values, these were extracted, otherwise where available the range of CD4 counts studied was recorded. The following measures of precision were extracted where available: %coefficient of variation for index and reference tests, no. of replicates per donor, no. of donors, CD4 count(s) of replicate sample(s). Where HIV positive data could be extracted, this alone has been included. Where this was not possible, combined HIV positive and negative data are included.

Correlation or linear regression data are inappropriate for evaluation of CD4 test accuracy and were not extracted. For this reasons, some studies which met the inclusion criteria had no data extracted.

7.1.3 Assessment of study quality

Study quality was assessed based on criteria outlined by the STARD (Standards for the Reporting of Diagnostic accuracy studies) initiative\(^4^8\). Particular focus was given to the following criteria:

-Did the study report the training received by the technicians prior to performing the index and reference tests?

-Did the study report time between blood draw and testing?

-Did the study use correlation analysis alone to report accuracy?

-Did the study report the distribution/range of CD4 of the sampled population?

-Were technicians performing index/reference tests blinded to the results of the other method?

7.2 Results

43 studies were accepted for inclusion in the review\(^1^4^3\). At least one published performance evaluation study was found for each of the following technologies: Coulter cytospheres, Dynal dynabeads, Guava Easy CD4, Partec CyFlow, BD FACSCOUNT, singleplatform flow cytometry (SPFC) using TruCount beads (Becton Dickinson), SPFC using Flow-Count fluorospheres (Beckman Coulter), SPFC using Perfect Count beads (Cytognos), SPFC using flow rate calibration, dual platform panleucogating (PLG), single platform PLG using Beckman Coulter FlowCARE, Sysmex pocH-100i with Dynal dynabeads, i+MED CD4 Select. No published, peer-reviewed performance evaluations of PointCareNOW or Guava Auto CD4 were found by our literature search. The studies took place in a wide range of resource-rich and resource-poor countries; the majority of studies took place in research, reference or teaching hospital laboratories.
7.2.1 Accuracy

7.2.1.1 Bias/mean difference

74% of studies reported bias, mean difference or median difference. 58% of studies performed Bland-Altman analyses with limits of agreement. 37% of studies reported bias for subgroups of the CD4 count range, e.g. for those samples <500, <350 or <200.

7.2.1.2 Misclassification probabilities

30% of studies reported misclassification probabilities or provided data from which misclassification probabilities could be calculated. Of these, the majority reported misclassification or sensitivity/specificity around a CD4 cut-off of 200 cells/µl, with only 3 studies providing data around a cut-off of 350 cells/µl. Misclassification probability data using a cut off of 350 cells/µl were available for the following technologies: Coulter cytospheres, Dynal dynabeads, Guava Easy CD4, Partec CyFlow.

Misclassification data are summarised in figures 1 and 2 below.

Both misclassification up and down (particularly down) are likely to be underestimates as none of the studies are restricted to the most clinically relevant range and many included HIV negative patients with high CD4 counts. Lutwana et al\(^6\) conducted a large study of manual technologies in Uganda; they recruited only HIV positive patients, the majority of whom had counts within the clinically important range (range in study: 0-900). Of samples with ‘true’ counts <350 cells/µl, Coulter cytospheres misclassified 1% as >350 cells/µl. However, of those with ‘true’ counts >350 cells/µl, 68% were misclassified as <350 cells/µl (indicating that 68% of patients not qualifying for treatment using flow cytometry would have done so if cytospheres were used). Dynal dynabeads had upward and downward misclassification probabilities of 6% and 30% respectively. Upward misclassification by either manual technology is however likely to be an underestimate as the majority of counts were very low (less than 25% of samples had counts >200 cells/µl); if the tests were to be used in a population with counts closer to the treatment threshold (as would be likely if used primarily in asymptomatic patients), upwards misclassification would be expected to be higher.

Karcher et al\(^9\) compared Coulter cytospheres and the Partec CyFlow Counter (both under field conditions), with dual-platform flow cytometry. Of samples with counts <350 cells/µl by flow cytometry, 29% were misclassified as having >350 cells/µl by the CyFlow Counter (indicating that 29% of patients eligible for treatment on the basis of a flow cytometry CD4 count may have been denied treatment if CyFlow-determined CD4 counts were used as the only criteria for treatment); downward misclassification probability was 7%. Coulter cytospheres in this study had upward and downward misclassification probabilities of 16% and 20% respectively.

Thakar et al\(^9\) compared Guava Easy CD4 with flow cytometry, and found 0% upward misclassification and 18% downward misclassification using a threshold of 350 cells/µl. The sample size was small however; calculating 95% confidence intervals gives confidence limits of 0-6% for upward misclassification and 6-38% for downward misclassification.

Importantly, we cannot tell from these misclassification probabilities the magnitude of the misclassification: are the samples that are being misclassified restricted to those whose ‘true’ values lie close to the threshold (e.g. within 10 CD4 cells/µl of the treatment cut-off), or are samples with
‘true’ values substantially greater or less that the treatment cut off also being misclassified? A visual impression of this can sometimes be gleaned from observing data plots, but a more useful measure would involve calculating the probability of these high-magnitude misclassifications.
Figure 1a. Misclassification up, using a CD4 cut-off of 200 cells/ul
Each bar represents results of one study. Height of bar represents % of samples misclassified as >200 cells/ul by the test under evaluation (i.e. for these samples, the index test gave a result >200, but the reference test gave a result of <200)

Figure 1b. Misclassification down, using a CD4 cut-off of 200 cells/ul
Each bar represents results of one study. Height of bar represents % of samples misclassified as <200 cells/ul by the test under evaluation (i.e. the index test gave a result <200, but the reference test gave a result of >200)
Figure 2a. Misclassification up, using a CD4 cut-off of 350 cells/ul
Each bar represents results of one study. Height of bar represents % of samples misclassified as >350 cells/ul by the test under evaluation (i.e. the index test gave a result >350, but the reference test gave a result of <350)

Figure 2b. Misclassification down, using a CD4 cut-off of 350 cells/ul
Each bar represents results of one study. Height of bar represents % of samples misclassified as <350 cells/ul by the test under evaluation (i.e. the index test gave a result <350, but the reference test gave a result of >350)
7.2.2 Precision

7.2.2.1 Within-laboratory precision

51% of studies reported within-laboratory precision data. These are presented in figure 3a below:

Figure 3a. Within-laboratory variability. %CV= coefficient of variation, NC=not clear, CC=Coulter cytospheres, DD=Dynal dynabeads

Gernow et al\(^4\) studied the reproducibility of Coulter cytospheres and found poor precision, with a coefficient of variation of 58%. The study by Landay et al\(^3\) however found precision levels more in keeping with the other technologies. The reason for this disparity of results is not clear. The CD4 counts of the 19 samples used for replicate analysis by Gernow et al were not stated, however they included 12 HIV negative samples which might be assumed to have high counts. Given that several papers found cytospheres to have poorer performance at higher counts, this may be in keeping with poor reproducibility in these replicates. However, Landay et al found better precision on a sample with a CD4 count of 1200 (%CV 3.5%) than on a sample with a CD4 count of 200 (%CV 10.8%). It should be noted that the manual methods, although employing simple technology, are labour intensive and require significant user skill. Inadequate training, lack of supervision or user fatigue may lead to poor performance of these techniques.
7.2.2.2 Between-laboratory precision

Six studies reported between-laboratory precision. Two were studies evaluating BD FACSCount, two were evaluating panleucogating, and two were studies evaluating bead-based single platform technology (BD TruCount beads and BC FlowCount fluorospheres). Four of the six studies also reported between-laboratory precision data for the reference test (usually a dual platform method).

These data are presented in figure 3b:

![Figure 3b Between-laboratory variability](image)

**Figure 3b Between-laboratory variability.** %CV= coefficient of variation, NC=not clear, DPFC=dual platform flow cytometry, DP PLG=dual platform panleucogating, BD=Becton Dickinson, BC=Beckman Coulter. Ref. tests=reference tests

Single platform flow cytometry using TruCount beads or FlowCount fluorospheres showed significantly less inter-laboratory variability than the dual platform comparators in these studies. In addition, Denny et al demonstrated significantly improved inter-laboratory precision using dual platform panleucogating compared with predicate technologies (which included dual or single platform conventional flow cytometry).

7.2.3 Conclusions of systematic review

This review highlighted the difficulties of answering clinically relevant questions about CD4 test performance from the published literature. The studies reviewed frequently concluded that a test had acceptable performance based only on correlation analysis, or on the demonstration of low levels of estimated bias over a wide range of CD4 counts, neither of which are reliable measures of
accuracy in this context. A minority reported clinically useful measures of accuracy, and few were carried out under field conditions.

The available data do however support the following statements:

1. Between-laboratory precision is likely to be superior with single-platform technologies (using BD TruCount beads or BC FlowCount fluorospheres) or panleucogating than with conventional dual-platform technologies
2. Manual methods, particularly Coulter cytospheres, may be associated with higher rates of misclassification and worse precision than flow-based methods, and are difficult to quality control

It is difficult to draw conclusions about the performance of Guava Easy CD4 or the Partec CyFlow instruments due to a paucity of appropriate data and, for the Partec instruments, a wide range of instruments, reagents and parameters used in different studies making it difficult to draw generalisable conclusions.

No published peer-reviewed literature were found relating to either Guava Auto CD4/CD% (now replacing Guava Easy CD4), or to the PointCareNOW instrument, and only one, non-independent study of the CD4 select assay was identified, which lacked precision or misclassification data.

Despite a lack of misclassification data on BD FACSCount, within-laboratory and between-laboratory precision were good and these findings are supported by EQA data.

It can be seen that whatever technology is chosen, there is variability associated with CD4 measurement. Performance characteristics vary between technologies, and these characteristics have important implications both for individual patient management and for HIV treatment programmes. It is therefore essential to consider performance as well as operating characteristics when choosing technology. However, these data are not always available in the literature and, currently, evaluations are not robust or comprehensive enough to give a clear idea of the comparative merit of different technologies.

What constitutes ‘acceptable’ performance remains undefined and may vary between sites, depending on local factors such as the distribution of CD4 counts among asymptomatic patients, how often patients come back for repeat CD4 testing and the implications of potential overtreatment (e.g. cost, promotion of drug resistance). It is relatively straightforward for a site to decide which technology best fits their needs based purely on operating characteristics; it may be harder to decide what performance characteristics are acceptable, and harder still to obtain data on test performance to inform choice.

Given the potential for error, participation in EQA programmes and access to QC reagents are essential. In the next sections we will describe the importance of quality control, and outline a suggested study design for future evaluations of CD4 enumeration technology.
8.0 Importance of EQA programmes and QC

?DB to help write

9.0 Suggested study design for future evaluations of CD4 technology

10.0 What is on the horizon: new CD4 technologies

Recent research and development of CD4 technologies have focused on creating accurate and reliable point-of-care (POC) assays. A POC test for CD4 enumeration could greatly improve access to CD4 testing in low-income countries with limited laboratory capacity.

Inverness Medical has a POC CD4 test, the Pima CD4, which is already commercially available in select markets. The test uses a disposable cartridge containing sealed reagents and a portable analyzer to provide absolute T-cell counts in whole blood within 20 minutes. The technology is based on static image analysis and counting principles but unlike current assays employing the same technology, the Pima is portable and does not require extensive laboratory equipment or training. Furthermore, the assay requires only 25 μL of capillary blood and has all test reagents contained within the disposable cartridge. The reagents are dried and therefore do not require refrigeration and no manual sample processing is required to obtain a result (http://pimatest.com/en/home.html).

Although it is commercially available, there have not been any independent evaluations of the Pima comparing the POC test to an acceptable gold standard assay.

mBio, a division of the Precision Photonics Corporation, has developed a system based on planar waveguide technology that is being applied to CD4 enumeration and the simultaneous diagnosis of HIV and common co-infections. The system uses an integrated fluidic cartridge and a fluorescence imaging system to provide automated cell counting. The technology, core multiplexed waveguide technology, has received FDA testing; however, there is no indication of the future availability of a POC CD4 test (http://www.mbidx.com/).

Daktari Diagnostics is a third company involved in the manufacture of a POC CD4 test. The Daktari system will consist of a hand-held, battery powered instrument and disposable plastic cartridges. Each test consists of applying a drop of blood from a fingerstick to the cartridge (only 10 μL of blood is required), capping the cartridge, and inserting the cartridge into the instrument. An absolute CD4 cell count result will be obtained within approximately six minutes. The system runs off a rechargeable battery, and is designed to be usable by anyone, anywhere in the world. The system is based on research in microfluidics and electrical sensing and employs: 1) a process for performing the assay on a drop of untreated whole blood, eliminating the time consuming and complex pre-processing steps required by flow cytometers; and 2) a process for counting cells using simple electrical impedance measurements, eliminating the use of costly optical methods. (http://www.daktaridx.com/).
The CD4 Initiative was established in 2005 by an award from the Bill & Melinda Gates Foundation and is managed through Imperial College of Medicine in London, UK with the objective of developing a new, low-cost, rapid POC test for measuring CD4+ T-lymphocytes. The Initiative put forth a call for design proposals and current awards are held by both Zyomyx Inc and ImmunoSite Technologies (http://www1.imperial.ac.uk/medicine/about/divisions/departmentofmedicine/infectiousdiseases/infectious_diseases/cd4_initiative/).

Zyomyx is developing a CD4 gauge similar to a thermometer that will quantify CD4 T-lymphocytes from whole blood obtained through a finger-prick; however, the prototype is still under development.

In addition to POC tests for CD4 enumeration, dried blood spots (DBS) have been investigated as a possible alternative for CD4 testing in settings with limited access to laboratory services. Redd et al. designed a technique for estimating mature T-cell numbers by calculating the amount of rearranged T-cell receptor β genes from dried blood spots of HIV-infected individuals. This was done by genomic DNA amplification with Real-Time PCR. The variability in the results and the failure to detect lymphocytes that are not fully or close-to-fully mature, demand more research before this can be presented as a viable alternative to current methods. The potential of combining HIV viral load quantification and CD4 enumeration on one single DBS could be an easily implementable and affordable alternative to the current, expensive methodologies.
11.0 References


11. Lynen L; Teav S; Vereecken C; De Munter P; An S; Jacques G; Kestens L Validation of primary CD4 gating as an affordable strategy for absolute CD4 counting in Cambodia Journal of Acquired Immune Deficiency Syndromes: JAIDS, October 2006, vol./is. 43/2(179-85)


17. Pattanapanyasat K; Phuang-Ngern Y; Sukapirom K; Lerdwana S; Thepthai C; Tassaneetrithep B Comparison of 5 flow cytometric immunophenotyping systems for absolute CD4+ T-lymphocyte counts in HIV-1-infected patients living in resource-limited settings. Journal of Acquired Immune Deficiency Syndromes: JAIDS, December 2008, vol./is. 49/4(339-47)


20. Manasa J; Musabaike H; Masimirembwa C; Burke E; Luthy R; Mudzori J Evaluation of the Partec flow cytometer against the BD FACSCalibur system for monitoring immune responses of human immunodeficiency virus-infected patients in Zimbabwe. Clinical & Vaccine Immunology: CVI, March 2007, vol./is. 14/3(293-8)


27. Nicholson JK; Stein D; Mui T; Mack R; Hubbard M; Denny T Evaluation of a method for counting absolute numbers of cells with a flow cytometer. Clinical & Diagnostic Laboratory Immunology, May 1997, vol./is. 4/3(309-13)


36. Young NL et al. Clinical field site evaluation of the FACScount for absolute CD3+, CD3+ CD4+, and CD3+ CD8+ cell count determinations in Thailand. Clinical & Diagnostic Laboratory Immunology, November 1997, vol./is. 4/6(783-6)

37. Higgins J. et al Evaluation of a single-platform technology for lymphocyte immunophenotyping Clinical and Vaccine Immunology, October 2007, vol./is. 14/10(1342-1348)

38. Jeganathan S; Bansal M; Smith DE; Gold J Comparison of different methodologies for CD4 estimation in a clinical setting. *HIV Medicine*, April 2008, vol./is. 9/4(192-5)


41. Srithanawiboonchai K et al Novel low cost assay for the monitoring of CD4 counts in HIV infected individuals J Acquir Immune Defic Syndr 2008;47:135–139


47. Systematic Reviews. CRD's guidance for undertaking reviews in health care.Centre for Reviews and Dissemination, University of York, 2008
48. www.stard-statement.org


51. Malone, Joseph L.; Simms, Thomas E.; Gray, Gregory C.; Wagner, Kenneth F.; Burge, J. Robert; Burke, Donald S. Sources of Variability in Repeated T-Helper Lymphocyte Counts from Human Immunodeficiency Virus Type 1-Infected Patients: Total Lymphocyte Count Fluctuations and Diurnal Cycle Are Important Journal of Acquired Immune Deficiency Syndromes. 3(2):144-151, February 1990


58. World Health Organisation Rapid advice: antiretroviral therapy for HIV infection in adults and adolescents - November 2009


APPENDIX A: SEARCH STRATEGY AND SEARCH TERMS

Search strategy

Databases: MEDLINE and EMBASE


Contacting experts (advisory group)

Search terms


Subject headings to include “CD4 antigen”, “Antigens, CD4”, “CD4 lymphocyte count”, “CD4-positive T-lymphocytes”, “technology”, “methodology”, “technique”, “evaluation”, “clinical evaluation”, “economic evaluation”, “evaluation research”, “evaluation and follow up”, “instrument validation”, “validation process”, “validation study”, “evaluation studies as topic”, “validation studies”, “flow cytometry”

Date

1990-presentation

Language

English only

Publication type/status

Only published work in peer reviewed journals
433 titles +/- abstracts reviewed for relevance

345 excluded as not relevant to study question ie not evaluations of CD4 technology

88 articles identified as potentially relevant (ie evaluations of CD4 enumeration technology)

1 study excluded as did not use flow cytometry as reference standard

1 study excluded as paediatric only

1 study excluded as not English language

7 conference abstracts excluded

14 studies excluded as index test not commercially available at current time

1 study excluded as evaluating complex flow cytometer (FACSCanto)

4 studies excluded as EQA studies, not evaluations

5 studies excluded as not evaluations of new technology (ie old technology or minor modification of existing tech)

13 studies excluded as not evaluations of new technology or not evaluations of CD4 technology

1 study excluded as used outdated FC as reference method

4 other studies contributed by advisory group

1 excluded as not commercially available

3 studies accepted for inclusion

40 studies accepted for inclusion

43 studies accepted for inclusion