Monitoring patients with HIV disease

Matthew Helbert, Judy Breuer

From early in the human immunodeficiency virus (HIV) pandemic it was noted that a long subclinical latent phase preceded the appearance of opportunistic infections. The use of immune monitoring to prompt prophylactic agents considerably reduced morbidity caused by opportunistic infection in the 1980s. In the past few years, as new antiretroviral treatments emerged but were threatened by resistance, viral monitoring has been used to guide such treatment. Consensus guidelines on laboratory monitoring and treatment in HIV have been published on both sides of the Atlantic. In this article, we discuss the natural history of HIV infections, current immunological and virological tests, and recommendations for monitoring protocols.

Natural history of HIV infection
Efficient infection of T helper cells, macrophages, and dendritic cells by HIV requires expression of the CD4 molecule and chemokine receptors. In early infection, HIV is characteristically trophic for the CCR5 chemokine receptor on macrophages and dendritic cells, explaining why the 1% or so of the white population with a CCR5 polymorphism are relatively protected from HIV.

Early HIV infection is characterised by active viral replication, with high amounts of viral RNA, recoverable virus, or viral antigens such as p24 being found. Subsequently, viraemia falls because specific cytotoxic T cells kill infected cells, replication is inhibited by non-lytic means, or the availability of uninfected target cells is exhausted. Thus, an initial precipitous fall in CD4+ cell counts may be followed by partial recovery at this stage.

Although the amounts of virus in blood fall, replication continues in lymphoid tissue until a new steady state of viraemia is attained. The mean half life of the virus in the plasma is approximately two days, suggesting that the continual production of virus and target cells is matched by an equivalent rate of clearance of virus and cell death.

Consequences of viral mutation can disturb this equilibrium. HIV uses reverse transcriptase (RT) to translate its RNA genome to DNA, an error prone process characterised by the emergence of mutated strains. Drug resistance is one consequence. In addition, the emergence of mutated strains. Drug resistance is one consequence. Drug resistance is one consequence.

A third consequence of mutation is increased binding to the CXCR4 chemokine receptor, conferring increased tropism for CD4+ T cells. T cell destruction is upregulated through several mechanisms including syncytium formation in lymph nodes, apoptosis, or a combination of these processes. HIV renders thymic and peripheral mechanisms for compensatory T cell regeneration dysfunctional. This model of pathogenesis is an oversimplification, but the net result in vivo is declining CD4 numbers and increasing viral replication. Production of up to 10^11 HIV particles and loss of 10^8 CD4+ T cells each day is typical.

Immune monitoring
CD4 COUNTS
The emergence of HIV in the 1980s coincided with advances in flow cytometry and monoclonal antibody production, enabling CD4+ T cell counts to become the key immune monitoring tool. The absolute CD4 count is more widely used than the CD4 percentage (percentage of lymphocytes expressing CD4) or the CD4 : CD8 ratio because most natural history studies (which have suggested thresholds for chemoprophylaxis) and drug trials have used the former. In addition, the CD4 : CD8 ratio may be perturbed by physiological expansions of the CD8+ population—for example, in response to HIV itself.

Flow cytometers enumerate cell populations on the basis of size (forward scatter), granularity (side scatter), and staining with up to four monoclonal antibodies, each conjugated to a differently coloured fluorochrome. Early technology required the operator to select (“gate”) populations with the physical characteristics of lymphocytes, within which the percentage of CD4+ cells was counted. This percentage was then related to the absolute lymphocyte count derived from haematology counters to give the absolute CD4 count, the so called two platform approach. These two processes are each the source of potential error:

1. Manually derived lymphocyte gates might be contaminated by non-lymphoid cells and debris. How this is overcome depends on the flow cytometry technology available:
   - Double staining of CD3 with CD4 or CD8 allows the exclusion of CD3−CD4+ (monocytes) and CD3−CD8+ (natural killer cell) populations.
   - With two colour cytometry, gating on an initial sample for CD45 (leucocyte common antigen) and excluding CD14+ (monocyte marker) cells improves the purity of the subsequently counted lymphocytes, but requires extra processing.
   - With three colour cytometry, simultaneous gating on CD45 and side scatter allows determination of two markers of...
Monitoring patients with HIV disease

Entire process Check no unexpected deviations from previous counts

hours. If a haematology count is required, it the process (table 1).

Control procedures should monitor all parts of

should be processed within six hours. Quality

Numerous factors a

INTERPRETATION OF CD4 COUNTS

The "lymphosum" should be checked; that is, the percentage of CD4+, CD8+, CD19+ (B cell), and CD56+ (natural killer cell) cells = 100% ±10%. US guidelines suggest doing lymphosum on each occasion. It is acceptable to do this only on new patients and in case of queries

Daily use of commercial lyophilised cells with analysis of results. Cells from normal donors may show unexpected variations and should not be used

TRANSIENT INCREASES

Women in follicular phase of the menstrual cycle

Diurnal variation in normal donors, reciprocating serum cortisol

Physiological stress (for example, post operative, immediately after exercise," or during bereavement)"

Exogenous steroids

Cytotoxic drugs

Table 1 Quality control of CD4 counts

| Machine optics | Fluorescent beads |
| Antibody | Negative controls not required for antibodies producing discrete positive and negative peaks, such as CD3, CD4, and CD8 |
| Gating and staining | The "lymphosum" should be checked; that is, the percentage of CD4+, CD8+, CD19+ (B cell), and CD56+ (natural killer cell) cells = 100% ±10%. US guidelines suggest doing lymphosum on each occasion. It is acceptable to do this only on new patients and in case of queries |

Table 2 Factors affecting CD4 counts

| Transient reductions | Diurnal variation in normal donors, reciprocating serum cortisol |
| | Physiological stress (for example, post operative, immediately after exercise, or during bereavement) |
| | Exogenous steroids |
| | Cytotoxic drugs |
| Transient increases | Intercurrent infections (for example, urinary tract infection and influenza) |

Table 3 Prophylaxis guidelines for adults

| Infection | CD4 Threshold |
| | |
| PCP | CD4 count below 200 ×10^6/l |
| Toxoplasmosis | CD4 count below 100 ×10^6/l |
| MAI | CD4 count below 75 ×10^6/l |
| CMV* | CD4 count below 50-100 ×10^6/l |

*Not routinely followed in the UK.

CD4 counts in newborns are higher than in adults and reach a peak at about 6 months of age (fig 1). In children with HIV infection, the CD4 count is a good predictor of outcome and

interest (for example, CD3 and either CD4 or CD8). The presence of anti-CD3 allows a consistency check for the total T cell count.

Four colour technology extends this approach by enabling CD45/side scatter gating, and collecting data on CD3, CD4, and CD8 from one tube. In addition, four colour counting eliminates tube to tube variation.

Computer selection of gates using these parameters can save operator time, but derived gates should be checked manually. For example, computer selected gates can exclude large granular lymphocytes.

(2) Separately derived lymphocyte counts require up to two extra steps (the use of a haematology analyser and differential white cell count) and this has been identified as a major source of error. Techniques such as precision fluids or incorporation of a known number of fluorescent particles for each volume of blood enable some flow cytometers to produce absolute counts directly and have become preferred methods. Switching from older technologies to these "single platform CD4 methods" might produce unexpected changes in absolute CD4 values, which require audit and liaison to ensure a smooth transition.

Single platform instruments using three or four colours might reduce running costs because they can perform single tube tests, do not require haematology counts, and have increased automation.

Samples for flow cytometry should be stored at 10–16°C and stained and fixed within 16 hours. If a haematology count is required, it should be processed within six hours. Quality control procedures should monitor all parts of the process (table 1).

Interpretation of CD4 counts

Numerous factors affect CD4 counts, as shown in table 2. For example, diurnal changes can affect the CD4 count by as much as 50%. It should be noted that some of the pathological states shown in table 2 can clinically mimic HIV infection. The consequences of these changes are:

(1) CD4 counts should be done at consistent times of day and avoided during acute illnesses.

(2) In the first few weeks after diagnosis, baseline values should be obtained. Subsequently, counts can be done every six months in asymptomatic patients or every three months after the onset of symptoms. More frequent testing is justified when treatment is being altered.

(3) CD4 counts cannot be used to diagnose HIV infection. Although anecdotal data and the author’s own observations suggest this does occur, substitution of CD4 counts for HIV serology is clinically invalid.

In the USA, the case definition for AIDS has been broadened to include all HIV infected patients with a CD4 count below 200 ×10^6/litre, for epidemiological purposes. In the UK, this practice was discontinued in early 1996, but at that time, prophylaxis was provided to patients with CD4 counts below 200 ×10^6/litre. Che- mor prophylaxis reduces the risk of pneumocystis pneumonia by up to 80% in patients with a CD4 count below this value and was the major reason for declining mortality in HIV infection before highly active antiretroviral regimens were available. Similar guidelines have been suggested for other prophylactic treatments (table 3).

CD4 counts in newborns are higher than in adults and reach a peak at about 6 months of age (fig 1). In children with HIV infection, the CD4 count is a good predictor of outcome and

Infection | CD4 Threshold |
| | |
| PCP | CD4 count below 200 ×10^6/l |
| Toxoplasmosis | CD4 count below 100 ×10^6/l |
| MAI | CD4 count below 75 ×10^6/l |
| CMV* | CD4 count below 50-100 ×10^6/l |

*Not routinely followed in the UK.
can be used in assessing whether antiretroviral agents are indicated, particularly when trends over time are available. CD4 based guidelines for pneumocystis pneumonia prophylaxis in childhood have not been determined, although it is clear that adult values should not be used. Because many children with HIV infection will never have been exposed to some pathogens, primary prophylaxis might be indicated, regardless of CD4 counts.

Changes in CD4 counts have been widely used as end points in trials of antiretroviral agents, to accelerate results and pre-empt irreversible damage. However, CD4 counts do not always correlate with clinical outcome, possibly because CD4+ lymphocytes are not key players at all stages of infection, and because CD4 counts do not reflect function of T cells. Viral load is a better measure of antiviral efficacy (see below).

Highly active antiretroviral therapy (HAART) reduces viral loads and generally improves CD4 counts. However, individuals on HAART can develop opportunistic infections after CD4 counts have risen above the thresholds shown in table 3. For example, there is a continued risk of cytomegalovirus (CMV) after the CD4 count has climbed above 100 × 10^6/litre, possibly because HIV irreversibly deletes specific T cell clones. Such opportunistic infections tend to occur in the first three months of HAART treatment and present atypically. Alternatives, such as monitoring with CMV polymerase chain reaction (PCR)—for example, only give incomplete information on timing of CMV prophylaxis.

To overcome this type of problem, either new CD4 count thresholds for chemoprophylaxis on HAART, or tests of responsiveness to specific pathogens are required.

Other phenotypic changes have been investigated as surrogate markers in HIV infection. For example, the expression of CD38, a marker of cytotoxicity, is increased on CD8+ lymphocytes in HIV infection. CD38 expression is associated with a poor prognosis, independent of CD4 count, especially when the number of CD38 molecules on each CD8 cell is estimated to be high. CD38 expression returns towards normal during antiretroviral treatment, but the use of CD38 has not been widespread however.

**Virological monitoring**

**Viral load**

In 1995, it was shown that the measurement of viral load by quantitative reverse transcriptase polymerase chain reaction (RT–PCR) predicted disease outcome from within six months of seroconversion, with viral loads of greater than 10^5 copies/ml (referred to as log_{10} 5.0) associated with a 10-fold increased risk of progression over five years. Further analysis has shown that, in patients with normal CD4 counts, even a single viral load measurement can predict outcome over the subsequent 10 years. In terms of long term forecasting, viral load produces better results than the immunological approaches mentioned earlier. In contrast, CD4 counts are better predictors of short term risk.

Coming together as it did with new insights into the pathogenesis of HIV infection and the advent of the more effective drugs, the notion that retarding the course of disease might relate to a reduction in viral replication was attractive. This hypothesis has been borne out by numerous studies. In patients receiving HAART, viral load is the single best predictor of survival, with maintenance of virus load below 10^3 copies/ml associated with slower progression to AIDS. Thus, viral load has been incorporated into guidelines for initiation and monitoring of antiretroviral treatment.

**Viral load assays**

Currently, there are three commercially available methods for measuring viral RNA:

1. Quantitative RT–PCR. RT is used to convert viral RNA into a DNA copy, which is
amplified by conventional PCR. In one commonly used assay (Roche, Basel, Switzerland), the quantification of the starting RNA relies on an internal reference sequence of known concentration, which is competitively amplified. The assay works best on EDTA anticoagulated blood and has also been used for the detection of HIV-1 in CSF.

(2) Nucleic acid sequence based amplification: NASBA works by a similar principle. However, no reverse transcription is required because the target can be RNA or DNA and thermocyclers are not needed for amplification. Again, the quantification is achieved by coamplification of an internal standard. NASBA is suitable for the measurement of virus in many specimens including EDTA plasma, serum, and semen.

(3) Branched chain DNA (bDNA): bDNA differs from the other two in being non-enzymatic. Instead, the signal from a captured RNA target is amplified by sequential oligonucleotide hybridisation steps. EDTA samples are required for bDNA.

Initially the bDNA method detected a wider diversity of strains but was less sensitive than the other two. Modifications now enable all three methods to detect not only the predominant UK and North American B HIV-1 strain, but also African strains, including the new O subtype. Only one assay, the Abbott RT–PCR (Abbott, Abbots Park, Illinois, USA) will quantitate the O subtype. The value for viral load detected also differs between assays, which might relate to differences in assay calibration. Caution should be exercised when comparing results from one test with another or changing assays. Increasingly automation is being introduced to maintain consistency as numbers of samples increase. Currently, only the Roche RT–PCR assay is fully automated, although the Chiron bDNA assay (Chiron, Emeryville, California, USA) can be run on a semi-automated analyser.

In all three assays, it is recommended that plasma be separated within hours of leaving the patient (bDNA four hours, RT–PCR and NASBA six hours). However, there is some evidence that HIV RNA in unseparated EDTA samples might be stable for up to 30 hours. For transportation to a remote laboratory, samples in centrifuge tubes in which a plug separates cells from plasma (for example, plasma preparation tubes; Becton Dickinson, Franklin Lakes, New Jersey, USA) are stable for up to five days at 4–25°C or for longer at −70°C. Common practice is to aliquot the plasma for storage to avoid the effects of repeated freezing and thawing.

Minimisation of traffic between pre-amplification and post-amplification areas and the use of plugged tips is fundamental to the avoidance of contamination. In-house controls should be included in every run, and Shewhart charts plotted to monitor intra-assay and inter-assay variation caused by reagent and operator related factors. A UK national external quality assurance scheme for viral load is under way.

INTERPRETATION OF VIRAL LOAD

Qualitative virological, rather than serological, detection might be required in some situations—for example, diagnosing seroconversion illness and screening blood products. Quantitative viral load assays are not suitable for screening or the diagnosis of HIV infection, because they might be less sensitive than qualitative tests, which detect proviral DNA.

During HIV infection, activation of the immune system upregulates HIV replication. Hence, transient increases in viral load might be expected after events such as vaccination. Such an effect has been observed after influenza vaccination, although this finding has not been reproduced. Increases in viral load have also been seen during the acute phases of tuberculosis in HIV infection, suggesting that, as in the case of CD4 counts, testing should not be done during acute illness.

Because a 0.3 log variation in viral load has been observed in untreated, stable patients, a sustained change of greater than 0.5 log is taken as reflecting a biologically relevant change in the degree of viral replication.

HAART consists of combinations of inhibitors of either RT (the nucleoside and non-nucleoside RT inhibitors) or HIV specific protease. Viral load monitoring helps clinicians by suggesting when HAART should be initiated and when it should be reviewed. Thus, all patients should have viral load measured every three to six months, whether they are on antiretroviral treatment or not.

There is general agreement that HAART should be started when viral load is detectable or above 10⁴ copies/ml. UK guidelines emphasise that treatment should be started if the viral load is above 5.5 × 10⁴ copies/ml and it should begin while the immune system is still intact—for example, when the CD4 count is above 350 × 10⁴/litre. Decisions about such early treatment should be balanced by considerations of the side effects of drug combinations that will be taken for the rest of the patient’s life.

After the introduction of HAART, the rate of viral load decline and the nadir are related to the durability of virological response. Viral loads of less than 5000 copies/ml of plasma are associated with a significant increase in life expectancy, reflected in the declining mortality from AIDS since the advent of HAART in 1995. In 40–50% of cases, the patient’s viral load falls to below 50 copies/ml within eight to 12 weeks of the onset of HAART and is accompanied, in 90% of patients, by a recovery in CD4 counts. A 2 log drop in viral load within the first two weeks predicts a lower risk of the subsequent emergence of viral resistance. The advent of ultrasensitive assays detecting less than 20–50 copies/ml has allowed even tighter control of viral replication, and patients whose viral load falls below 20 copies/ml on treatment have been shown to be at lower risk of virological treatment failure.
However, 20–30% of patients do not show a reduction in their viral loads after HAART, although about 35% of these experience an improvement in CD4 counts. Such primary drug failures might reflect poor compliance or absorption—for example—or viral resistance, which is most often acquired during treatment. The measurement of blood drug concentrations can exclude the first two causes of relapse, and is available in the UK on a referral basis (for more information see http://liv.ac.uk/hivgroup). In other patients who have failed treatment, viral resistance testing has been shown to inform the choice of subsequent treatment and improve patient outcome.

To detect all kinds of drug failure, viral load should be measured three to four weeks after initiating treatment and every three to six months thereafter.

HIV persisting in long lived T cells, macrophages, and the brain is a source of emergent drug resistance if replication is allowed to resume. This might be signalled by a rise in the viral load, although there is little agreement on what represents virological relapse. One view is that treatment should be altered if viral load increases by 0.5 log above the nadir, but a more conservative view is that viral load should be allowed to reach pretreatment values before a change of drugs. However, this latter approach might allow further accumulation of resistance mutations (see below).

MONITORING FOR ANTIVIRAL RESISTANCE

Resistance to antiretroviral drugs occurs as a result of mutations in the POL gene, which encodes RT and RNA polymerase. Each drug tends to induce its own signature set of mutations, which reduce the activity of that particular agent. For example, resistance to the nucleoside RT inhibitor zidovudine is caused by combinations of mutations of up to six discrete amino acid positions in RT. Some mutations confer cross resistance within the nucleoside RT inhibitor or protease inhibitor groups, whereas others that cause resistance to one drug increase sensitivity to another.

Drug resistance is detected by either genotypic or phenotypic assays:

1. Point mutation assays detect specific mutations using specific PCR and differential hybridisation approaches. These methods are the simplest and cheapest, although impractical when many different parts of the genome need to be analysed or undescribed mutations are being sought.

2. DNA sequencing assays rely on dideoxy chain terminator or gene chip technologies to sequence the POL gene and detect both known and novel resistance mutations. Derived sequences are compared with those of wild-type virus to pinpoint mutations. This process is costly and not entirely reliable because wild-type virus might be polymorphic, and mutations in POL do not always correlate with phenotypic resistance. Both types of genotypic assays have limited sensitivity for minority resistant strains, require the patient to be taking the drug of interest (resistant strains might become undetectable within 48 h of stopping the drug), and require a reasonable amount of viral breakthrough (for example, two to 5000 copies/ml).

3. Phenotypic assays rely on culturing virus, which is costly and might select outgrowth of atypical strains. The plaque reduction assays detect foci of cytopathic effects in a monolayer of CD4+ HeLa cells. Growth inhibition assays measure p24 production by cocultures of viral isolates and stimulated fresh lymphocytes. The concentration of drug required to inhibit 50% of cytopathic effects (IC50) or p24 production is calculated.

The recombinant viral assay combines genetic and phenotypic approaches. Patient virus POL genes are amplified and cloned into a replication deficient POL deleted proviral virus or a vector containing a reporter gene. The recombinant virus can then be assayed phenotypically for resistance. These assays are particularly suitable for detecting minority species and previously undescribed mutations. However, it is not yet known whether the potential advantages of this assay will translate into tangible clinical benefits.

Baseline phenotypic and genotypic resistance are both highly predictive of the subsequent response to non-nucleoside RT inhibitors or protease inhibitors. Data from a randomised trial show improved virological outcomes when treatment switches are tailored to individual resistance testing rather than rigid guidelines, although correlations between genotypic resistance and improvements after drug switching have not always been seen.

Although there are currently no consensus guidelines on resistance testing, it might be appropriate to test high risk patients, including those requiring salvage treatment and pregnant women, to minimise the use of inappropriate drugs. Resistance testing is also being offered by some to patients receiving post exposure prophylaxis after an inoculation injury to optimise treatment.

Conclusions

There is little doubt that monitoring tests have contributed to the improvements in HIV morbidity seen as a result of chemotherapy and effective antiretroviral treatment. These are summarised in table 4. The very high costs of HAART are partly defrayed by decreased inpatient stays, and the use of these drugs can be rationalised further by very sensitive viral load assays and reliable CD4 counts. Improved testing for viral resistance might allow even more fine tuning of treatment protocols, but unless the costs for genotyping fall, this is unlikely to become a routine part of monitoring. Assays for functional reconstitution are needed to guide the use of chemophrophylaxis in patients taking HAART. Follow up of CD4 counts and the risk of opportunistic infection in patients on HAART.
remains an area of active research. Together, these tests are likely to contribute to more frequent and sustained improvements for patients with HIV.

The high technology tools described above are unlikely to be useful for most people with HIV living in the developing countries. For these 14 million or so individuals, access to viral monitoring or antiretroviral drugs is likely to be very limited. As was the case in the 1980s in the developed world, prophylaxis and treatment of opportunistic infections (especially tuberculosis) is a more immediate alternative. Despite this, there are very few peer reviewed assessments of “appropriate technology” for immune monitoring in HIV infection. For example, a manual technique for CD4 counts appeared to perform well in a preliminary study on Ugandan samples and β, microglobulin and neopterin showed some value in the assessment of HIV infected Romanian orphans. Until such approaches are validated, laboratory monitoring of HIV infection in the developing world remains out of reach.

We would like to thank Jane Norman and Lorna Miller for sharing their wisdom and experience with us.

Table 4 Monitoring in HIV infection: summary points

<table>
<thead>
<tr>
<th>Test</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 counts</td>
<td>Predict immediate risk of opportunistic infection and indicate use of prophylaxis</td>
</tr>
<tr>
<td>Viral load</td>
<td>Predicts the long term risk of disease progression and indicates the use of HAART</td>
</tr>
<tr>
<td>State of the art testing</td>
<td>New assays are required for monitoring immune improvement on HAART</td>
</tr>
</tbody>
</table>

HAART, highly active antiretroviral therapy.


