Remote sensing of HIV care programmes using centrally collected laboratory results: Can we monitor ART programme effectiveness?

Carl Morrow, Alison Munro, Marisa Wilke, Ruth Stark, Robin Wood

Aims. We describe a monitoring system at population level of patients on antiretroviral therapy (ART) using centrally collected laboratory data. We demonstrate an analogous process of remote sensing using a large set of laboratory results and illustrate the tremendous density of information stored. We moved from an individual to a community view of ART rollout, similar to remote sensing used in the earth and biological sciences when the spatial scale of the investigation is too large to be performed at ground level.

Methods. This was a retrospective cohort study of patients from January 2004 to June 2011. A total of 188 759 individual laboratory results representing 26 445 patients were analysed for average CD4 and viral load by year.

Results. The data showed an increasing state of health of the population and allowed for hypothesis generation when the trends did not follow expected paths.

Conclusion. In this analysis we moved away from individual-centred data to population-level data in order to assess ART programme performance. Routine patient-monitoring data had great utility in assessment of population health. These methods are useful in monitoring and evaluation and effectiveness studies as they are easy to collect, reliable (not requiring much human matching or interventions) and scalable from a single clinic to an entire population. The larger the sample size, the more reliable the results, as confounders (such as incorrectly identified transfers out, lost-to-follow-up patients and transfers in) would be removed.

S Afr Med J 2012;102(6):501-505.

Since the introduction of treatment for the human immunodeficiency virus (HIV) epidemic through the large-scale rollout of antiretroviral therapy (ART), the population of people receiving therapy has increased dramatically.[1] As a result, it is becoming increasingly difficult and expensive to collect and collate data from individual patients in order to compile reports of patient entry and outcomes for insight into rollout programme success.[2,3] The need for continued improvement in monitoring and evaluation (M&E) systems was again highlighted in the National Strategic Plan for HIV and AIDS, STIs and TB, 2012 - 2016 (NSP).[4] Furthermore, the critical role of integrating M&E with electronic systems has been emphasised by the World Health Organization (WHO).[5]

ART drugs are highly effective at suppressing HIV in humans, as demonstrated by randomised controlled clinical trials.[6] However, the controlled environment within trials is very different from the clinical setting where the impact of ART on a population is determined primarily by programmatic issues of treatment availability, accessibility and delivery.[7] An innovative approach to monitoring ART programmes is required that is simple and efficient, and provides results that can be reliably interpreted, in settings such as South Africa (SA), where drug delivery occurs in an overburdened health system to sick, frequently impoverished individuals and adherence to treatment regimens cannot be assumed.

Extensive use is made of remote sensing in the earth and biological sciences, when the spatial scale of the investigation is too large to be done at ground level on individual plants or geographical units.[8] Essentially an image is taken of the landscape from a distance and the information contained in the image is analysed in order to interpret patterns that are present. In these examples, technology is used to assess the health of forests and other landscapes, monitor nutrient levels in standing farm crops and identify early outbreaks of disease.[8] This information is then used to target specific interventions.

Currently routine laboratory tests are seen as a necessary expense in individual patient care.[9] The records are stored in the individual's clinical notes but also, importantly, in centralised databases maintained by clinical laboratory services. The aim of this study was to demonstrate an analogous process of remote sensing by making use of a large, well-collected data set to illustrate the tremendous density of stored information that can be used to gain valuable insights into programme performance and population responses to ARV treatment.

Methods

This was a retrospective cohort study of patients within a SA non-government organisation (NGO)-run clinic system for the period January 2004 to June 2011. Nineteen data sets, representing 17 clinics from 5 of the 9 provinces in SA, were supplied from the Southern African Catholic Bishops' Conference/Catholic Relief Services (SACBC/CRS) ART programme.

Each data set consisted of 3 comma-separated variable files (extension .csv) containing data of laboratory tests, a main, anonymised, clinical and demographic data file and regimen information. Individual files were imported into a Microsoft Access database. Duplicate records, records with date errors and

Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, Department of Medicine, University of Cape Town
Carl Morrow, PhD
Robin Wood, DTM&H, MMed, FCP (SA)

Southern African Catholic Bishops' Conference (SACBC) AIDS Office, Pretoria
Alison Munro

Catholic Relief Services South Africa, Johannesburg
Marisa Wilke, PhD
Ruth Stark, PhD

Corresponding author: *C Morrow (carl.morrow@hiv-research.org.za)*

mismatched dates (e.g. patient outcome before enrolment) were removed, resulting in 47 107 valid patient records, 34 907 regimen records and 410 352 blood tests comprising 141 019 blood sample records. A total of 28 186 patients were recorded as having started ART. Approval to analyse the data was granted by the institutional ethics review board (RecRef 169/2007).

HIV-positive men, women and children who were receiving ART and had one or more CD4 or viral load blood result were included in the analysis. The last viral load and CD4 prior to ART initiation (baseline) and all CD4 and viral load values while receiving ART (before the recorded outcome date) were extracted yielding a database of 188 759 individual laboratory results, representing 26 445 patients. No cut-off was applied to the period of time between baseline sample date and ART start date. Within this analysis there were 960 patients whose first CD4 or viral load measurement, or both, was more than 6 months prior to treatment start.

CD4 analysis

Using the patient identification number, the first CD4 result for each patient was ascertained, and the time to the following samples was calculated. These values were grouped into 6-month time periods to yield an average CD4 recovery for the study group. Each patient provides one point in each time period (depending on time receiving treatment). The mean CD4 for each patient in each year was determined. The means were then grouped into 100 unit CD4 strata. Each year was represented as a line with the CD4 category along the *x*-axis and the number of patients on the *y*-axis. Alternatively the strata were divided into the biologically relevant categories: 0 - 200, 200 - 350, 350 - 500 and >500 CD4 cells/mm^3. Each calendar year was plotted as a bar with the proportion of the population in each CD4 category comprising the components. It should be noted that if a patient had been receiving treatment for more than 1 year, their results would appear in each of the years receiving treatment.

Viral load analysis

The viral load trajectory for each patient over time was determined by identifying the first viral load (baseline) and then grouping the viral loads into 6-month time periods and presenting the proportion of the group with viral load in the different categories: suppressed, with a viral load of <50, 50 - 400 and >400 copies/ml.

The mean $\log_{10}$ viral load for each patient for each year was calculated and the proportion of the population in each of the one log unit categories was plotted against the calendar year. It should be noted that if a patient had been receiving treatment for longer than 1 year, results would appear in each of the years of treatment.

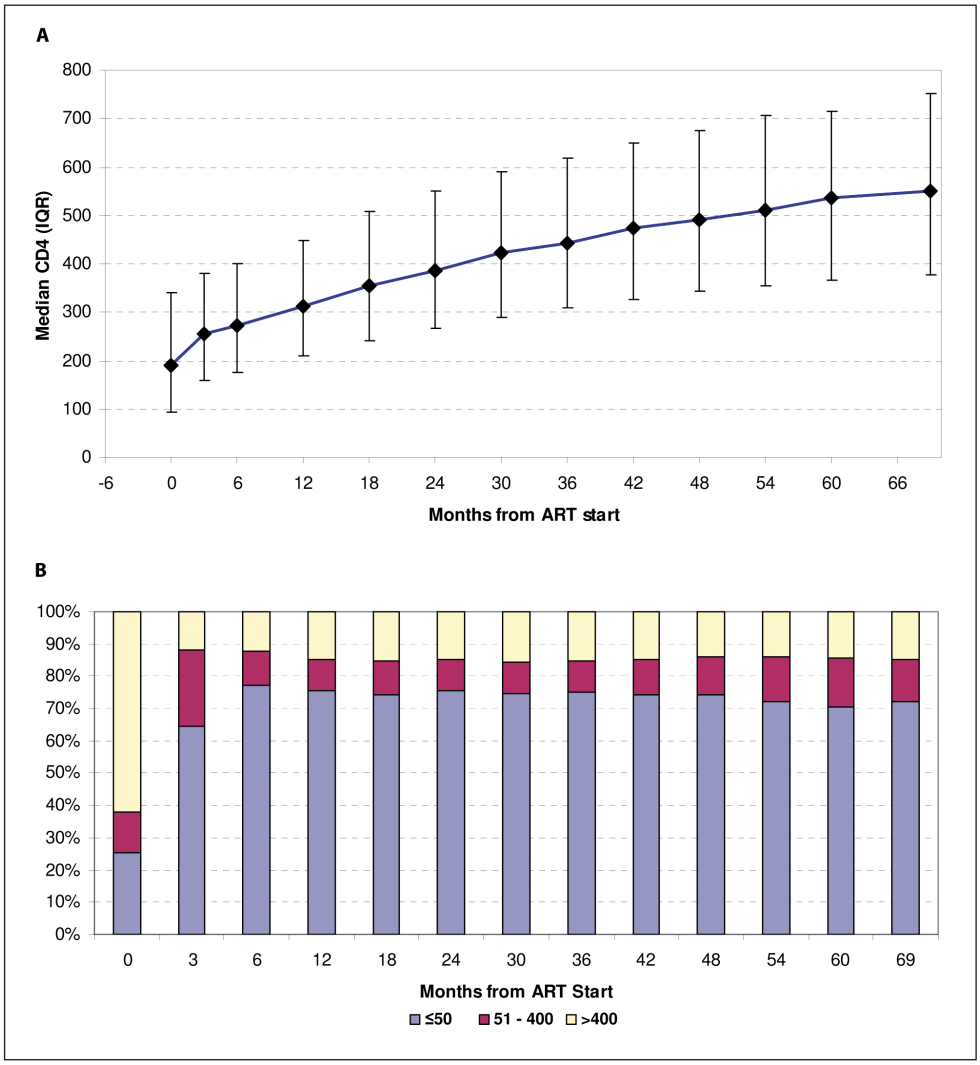

Fig. 1. Patient recovery. (A) Median CD4 and interquartile range (IQR) (cells/mm^3). (B) Proportions of viral load suppression over time showing the efficacy of the antiretroviral therapy in this clinic system.

Results

When presenting data on individual patient recovery, it is common to see curves showing the increase in CD4 value over time (Fig. 1A), confirming drug efficacy in people who are adherent to treatment. The error bars demonstrate diversity in recovery, with treatment failures and non-adherence influencing the median CD4 value. It should also be noted that there are fewer and fewer patients represented as one moves to the right of the figure.

A summary of the mean CD4 value for each patient in each year, enabling an assessment of total clinic population health, is presented in Figs 2 and 3. In Fig. 2 the area under the curve represents the size of the treated population; health can be assessed by observing where the peak of the curve lies and by noting the bulge in later years which represents patients with CD4 counts >400 cells/mm^3. Categorisation of the mean CD4 value for each patient for each year into CD4 strata yielded Fig. 3. This very clearly shows the improvement in population health and thus the effectiveness of ART rollout.

This analysis did not take patient age into account;

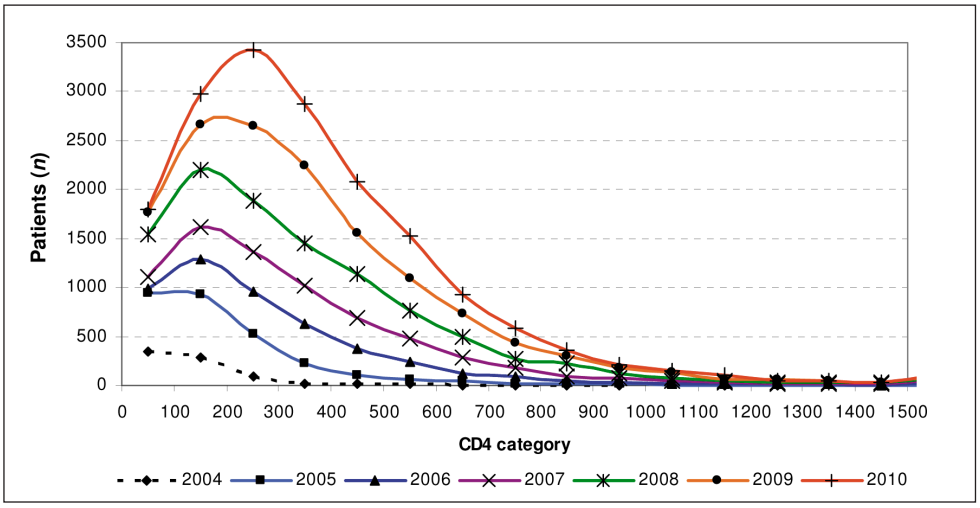

Fig. 2. Mean annual CD4 category (cells/mm³) showing increasing size of the patient population over time along with the improving health of the population demonstrated by the peak moving to the left along with the increasing proportion of the curve to the left of the peak.

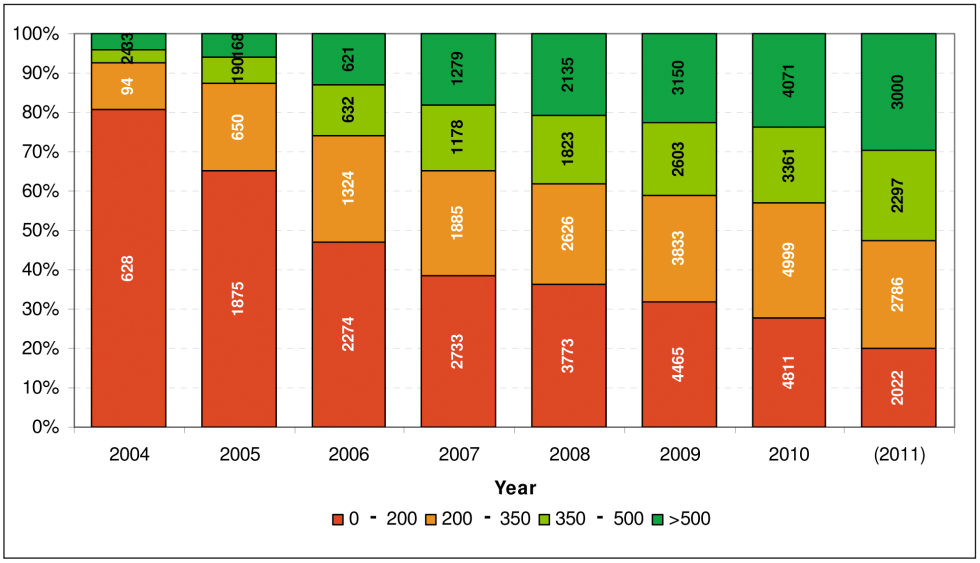

Fig. 3. Mean annual CD4 (cells/mm³) divided in strata; the improving health of the population over time is seen by the diminishing size of the lower categories. The values in the boxes are the number of patients in the category. The 2011 data are from January to June only.

therefore, it could be argued that the increases seen are due to an influx of children into the clinics. However, Table 1 shows that the proportion of children (under 10 years of age) is small, and has been decreasing over the years, so that child data have not influenced the trends towards higher CD4 values.

The traditional view of the improvement in response to treatment of individual patient viral loads over time is shown in Fig. 1B. We would expect all ART-naïve patients to have had a viral load >50 copies/ml. The conclusion here is that over the full monitoring period about 25% of patients are transferring into the clinics and are virally suppressed and currently receiving ART treatment. This is significant from a monitoring and evaluation perspective as this overview would be difficult to capture in a clinic setting. In moving away from a patient-centred to a community-centred view it is observed that the population viral suppression rate has increased dramatically over the 7-year study period, with over 60% of the population having an average viral load of <400 copies/ml in 2010 (the last complete year in the analysis).

Discussion

We clearly demonstrated that laboratory data can be used to view a care system from a distance to achieve remote sensing of the population's health. Once data were compiled, a relatively simple analysis gave a very powerful overview of the population level response to ART which is of great importance in the assessment of the impact of public health ART rollout. This is very different to the traditionally held opinion that the patient folder is the primary source of information needed to assess treatment programme efficacy.

There are many advantages to using these laboratory data in this evaluation role: data are collected at a more centralised location (a single pathology laboratory servicing a region) and can be accessed as a single download as opposed to collating monthly, or quarterly, records into a continuous history; results are reliable as they are directly used in clinical practice and so internal quality controls are already in place to ensure consistency and accuracy of the testing.

Minimal change is needed to the existing health infrastructure in order to implement this type of analysis. It can be started immediately and existing historical data allow for comparisons of current clinical systems with those of the past, without having to wait for the accumulation of new indicators. There is a minimal cost as all data are routinely collected; all that is needed is the time required for analysis. With some development, a reporting function can be built into existing databases to generate these results automatically.

We report a very 'clean' analysis to show an ideal, well-run clinic situation. Unselected, real-world data are more chaotic but can still be analysed in this way. Not taking the patient's ART start date into account resulted in a larger data set (222 901 records) but gave equivalent aggregate CD4 and viral load curves. These are not shown because the CD4 and viral load recovery trajectories of the complete data are not comparable with the traditional recovery curves (Figs 1A and 1B). The pretreatment laboratory data show an initial decline in CD4 after programme entry, followed by a sharp rise with ART initiation. Mirroring this is a steady increase in viral load followed by a sharp decline. These data cannot be used reliably because of the selection bias in this group with the CD4 threshold and other clinical data governing when the patients initiate ART. Many clinics make use of electronic pharmacy systems[10] that could

Table 1. Number of laboratory samples attributed to children aged <10 years, showing the decreasing proportion of samples over time

Year	Total samples (*N*)	Age ≤10 years in year of sample (*N*)	Age >10 years in year of sample (*N*)	Percentage of ≤10-year-olds (%)
2004	1 645	164	1 481	9.97
2005	8 101	517	7 584	6.38
2006	14 472	946	13 526	6.54
2007	21 498	1 384	20 114	6.44
2008	30 306	1 641	28 665	5.41
2009	40 606	1 992	38 614	4.91
2010	52 070	2 424	49 646	4.66
Overall	168 698	9 068	159 630	5.68

be linked to the laboratory data to give an accurate start date for the analysis.

We have performed this analysis on other data sets of different data quality (RW and CM, unpublished data) and have found that the curves seen in Fig. 2 can move 'backwards' with the peak in later sampling years at a lower CD4 value and the proportion of patients in higher CD4 categories becoming less. If the population size is still increasing, this could be attributed to the fact that the clinics are recruiting patients who are sicker, or that existing patients are failing treatment. If the population is decreasing, then the conclusion could be that the clinic is losing healthy patients and retaining or recruiting sick ones. It can be seen that these analyses allow for hypothesis generation that can then be investigated within the clinic or district. In this situation, Fig. 3 would show 'flat' strata with little increase in the proportion of patients in the higher CD4 categories over time. It is not possible to elucidate the individual factors that may be contributing to a poor performance curve (dysfunctional clinic, poor adherence, cultural issues, associated issues such as alcohol or drug abuse, etc.), but a broad-scale analysis can be made that rapidly identifies areas of concern to which investigations can then be directed.

The minimum requirements for this analysis are that the data are available in electronic format, and include sample date, sample type (CD4 or viral load), value, and patient identifiers to allow for per person analysis. The sample location would also be needed for clinic comparison or other spatial analyses.

Patient identifiers in the laboratory database are used to link records together. Although useful, they are not critical to these analyses; in the absence of identifiers, patients that have more than one test performed in a year would be double counted, and the area under the curve would represents the number of tests rather than number of patients. In our experience, linking different laboratory systems together is complicated, owing largely to the difficulty of linking patient identifiers. As a consequence, large-scale, centrally stored data are the most useful but do challenge the perspective of the current trend of devolving CD4 testing to the clinics, using point-of-care machines to close the gap between testing and reporting, which prevents the loss of patients.[11] Creation of a network of these machines would prevent loss of these data.

ART programme success relies on its effectiveness. As previously shown,[12,13] there is a need to assess the viability, successful implementation and uptake of treatment interventions in the public health setting following demonstration of the efficacy of a treatment in the artificial environment of a clinical trial or pilot study. Using the methods described here, it is possible for direct assessment of the impact of ART on the population at large. It should be kept in mind that only the 'on treatment' population is being monitored,[14] but, with increased rollout, the size of the population pool with uncontrolled HIV not having yet entered the ART programme will decrease.

We have shown that population CD4 values are increasing, which will result in comorbid diseases such as tuberculosis[15] decreasing in incidence. Mean population CD4 values in Africa have been summarised[16] and, as Fig. 2 shows, the distribution remains very much lower than in the uninfected population in spite of ART.

That population viral load is decreasing can be used as a surrogate measure of the transmission rates within the community.[14] For this reason we do not need to remove newly recruited patients as they are still a source of infection until they have become virally suppressed. The value in performing a baseline viral load measurement at treatment start is confirmed as this would permit health managers to assess the pool of virus in the community or health district for which they are responsible. In our analysis we calculated an average $\log_{10}$ viral load for each patient in the year and presented this as a distribution of values, while the community viral load (CVL) used the last viral load in each year.[14]

Monitoring the progress of a treated population is complex.[2] While it is relatively easy to collect data on programme entry from clinic-based records, it becomes very difficult to ascertain programme losses and cycling (multiple transfers in and out of clinics). With the widespread availability of treatment, an outcome of lost-to-follow-up is no longer equivalent to death.[17] Transfers among clinics are becoming increasingly common as patients optimise their care strategy;[18] this is complicated to monitor at a district level as assessment of an individual clinic's lost-to-follow-up patients may include those 'transferred out' to a nearby clinic but unrecorded, while new patients entering a clinic's programme may be unreported 'transfer ins'.[14] By looking at the programme entry viral load, an estimate of transferring in patients can be made and it is interesting that the 25.4% unrecognised transfers rate seen previously[18] is similar to the results we have seen based on laboratory records. Furthermore, as this analysis is designed to take a regional view of a programme, provided the scale was large enough (health district level), individual transfers would not influence the population level CD4 and viral load assessments.

Indicators of a patient's functional status are often recorded with ambiguity and misclassification.[2] Monitoring the CD4 status reveals the population level of health and, by association, predicted functional status. There are moves to redefine indicators to achieve better evaluation of programme and population coverage and of programme quality (e.g. PEPFAR[19]). Our analysis provides data to assess both of these and we show development of new indicators that offer an overview of programme effectiveness at population level.

Conclusion

In this analysis we moved away from individual-centred to population-level data in order to assess ART programme performance. We showed that routine patient-monitoring data had great utility in assessment of population health. The numbers of patients included in our analysis are approximately equivalent to those needing ART (40 000) within a typical health subdistrict of 400 000 - 500 000 residents.[20]

These methods are useful in monitoring and in evaluation and effectiveness surveys, as the data are easy to collect, reliable (not needing much human matching or interventions) and scalable from a single clinic to an entire population. Indeed, the larger the sample size, the more reliable the results will be as such confounders as incorrectly identified transfers out, lost-to-follow-up patients and transfers would be effectively removed.

Acknowledgements. RW is funded in part by the Cost-Effectiveness of Preventing AIDS Complications (CEPAC) funded by the National Institutes of Health (NIH, 5R01AI058736-02); International Epidemiologic Database to Evaluate Aids with a grant from the National Institute of Allergy and Infectious Diseases (NIAID: 5U01AI069924-02); USAID Right to Care (CA 674 A 00 08 0000 700) and the South African Centre for Epidemiological Modelling and Analysis (SACEMA).

References

1. WHO, UNICEF, UNAIDS. Global HIV/AIDS Response: Epidemic Update and Health Sector Progress Towards Universal Access: Progress Report 2011. Geneva: WHO, 2011:97. http://www.who.int/hiv/pub/progress_report2011/en/index.html (accessed 13 January 2012).
2. Hoskins S, Weller I, Jahn A, et al. An appraisal of indicators used to monitor the treated population in antiretroviral programmes in low-income countries. AIDS 2010;24(17):2603-2607. [http://dx.doi.org/10.1097%2FQAD.0b013e32833dd0d3]
3. Harries AD, Gomani P, Teck R, et al. Monitoring the response to antiretroviral therapy in resource-poor settings: the Malawi model. Trans R Soc Trop Med Hyg 2004;98(12):695-701. [http://dx.doi.org/10.1016%2Fj.trstmh.2004.05.002]
4. National Strategic Plan for HIV and AIDS, STIs and TB, 2012-2016: Draft Zero for Consultation. August 2011. http://www.sanac.org.za/files/uploaded/519_NSP%20Draft%20Zero%20110808%20pdf%20%20final.pdf (accessed 9 January 2012).
5. WHO. Global Health Sector Strategy on HIV/AIDS 2011-2015. Geneva: WHO, 2011. http://whqlibdoc.who.int/publications/2011/9789241501651_eng.pdf (accessed 13 January 2012).
6. Mathis S, Khanlari B, Pulido F. Effectiveness of protease inhibitor monotherapy versus combination antiretroviral maintenance therapy: a meta-analysis. PLoS One 2011;6(7):e22003. Epub 2011 Jul 19. [http://dx.doi.org/10.1371%2Fjournal.pone.0022003]
7. Lawn SD, Myer L, Wood R. Efficacy of antiretroviral therapy in resource-poor settings: are outcomes comparable to those in the developed world? Clin Infect Dis 2005;41(11):1683-1684 [http://dx.doi.org/10.1086%2F498030] [PMID: 16267745]
8. Navalgund Ranganath R, Jayaraman V, Roy PS. Remote sensing applications: An overview. Current Science 2007;93(12):1747-1766. (http://cs-test.ias.ac.in/cs/Downloads/article_42132.pdf)
9. Stover J, Bollinger L. The Costs of Antiretroviral Therapy in Low and Middle Income Countries: Present Commitments and Future Needs. Glastonbury, CT: Futures Institute, 2008. (http://www.futureartcosts.org/The_Costs_of_Anti-Retroviral_Therapy.pdf (accessed 18 November 2011).
10. Wood R, Kaplan R, Bekker LG, Brown S, Rivett U. The utility of pharmacy dispensing data for ART programme evaluation and early identification of patients lost to follow-up. S Afr J HIV Med 2008; Autumn:44-48. (http://www.ajol.info/index.php/sajhivm/article/viewFile/34866/6496)
11. Peter T, Badrichani A, Wu E, et al. Challenges in implementing CD4 testing in resource-limited settings. Cytometry B Clin Cytom 2008;74, Suppl 1:S123-130. [http://dx.doi.org/10.1002%2Fcyto.b.20416]
12. Coetzee D, Hilderbrand K, Boulle A, Draper B, Abdullah F, Goemaere E. Effectiveness of the first district-wide programme for the prevention of mother-to-child transmission of HIV in South Africa. Bull World Health Organ 2005;83(7):489-494.
13. Tchendjou P, Same-Ekobo C, Nga A. Effectiveness of multidrug antiretroviral regimens to prevent mother-to-child transmission of HIV-1 in routine public health services in Cameroon. PLoS One 2010;5(4):e10411. [http://dx.doi.org/10.1371%2Fjournal.pone.0010411]
14. Das M, Chu PL, Santos GM. Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco. PLoS One 2010;5(6):e11068. [http://dx.doi.org/10.1371%2Fjournal.pone.0011068]
15. Lawn SD, Myer L, Edwards D, Bekker LG, Wood R. Short-term and long-term risk of tuberculosis associated with CD4 cell recovery during antiretroviral therapy in South Africa. AIDS 2009;23(13):1717-1725. [http://dx.doi.org/10.1097%2FQAD.0b013e32832d3b6d]
16. Thakar MR, Abraham PR, Arora S. Establishment of reference CD4+ T cell values for adult Indian population. AIDS Res Ther 2011;8:35. [http://dx.doi.org/10.1186%2F1742-6405-8-35]
17. Dalal RP, Macphail C, Mqhayi M, et al. Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa. J Acquir Immune Defic Syndr 2008;47(1):101-107. [http://dx.doi.org/10.1097%2FQAI.0b013e31815b833a]
18. Government of Malawi Ministry of Health Quarterly HIV Programme Report on HIV Testing and Counselling, Prevention of Mother to Child Transmission and Antiretroviral Therapy: April - June 2011. http://www.hivunitmohmw.org/uploads/Main/Quarterly%20HIV%20Programme%20Report%202011%20Q2.pdf (accessed 18 November 2011).
19. The President's Emergency Plan for AIDS Relief (PEPFAR). Next generation indicators reference guide. http://www.pepfar.gov/documents/organization/81097.pdf (accessed 18 November 2011).
20. Statistics South Africa. StatsOnline. Pretoria: Statistics South Africa, 2010. http://www.statssa.gov.za/publications/statsdownload.asp?PPN=P0302&SCH=4696 (accessed 1 March 2011).

Accepted 13 February 2012.