Articles

Cost-effectiveness of different strategies to monitor adults on antiretroviral treatment: a combined analysis of three mathematical models

Daniel Keebler*, Paul Revill*, Scott Braithwaite, Andrew Phillips, Nello Blaser, Annick Borquez, Valentina Cambiano†, Andrea Ciaranello†, Janne Estill†, Richard Gray†, Andrew Hill†, Olivia Keiser†, Jason Kessler†, Nicolas A Menzies†, Kimberly A Nucifora†, Luisa Salazar Vizcaya†, Simon Walker†, Alex Welte†, Philippa Easterbrook, Meg Doherty, Gottfried Hirnschall, Timothy B Hallett

Summary

Background WHO's 2013 revisions to its Consolidated Guidelines on antiretroviral drugs recommend routine viral load monitoring, rather than clinical or immunological monitoring, as the preferred monitoring approach on the basis of clinical evidence. However, HIV programmes in resource-limited settings require guidance on the most cost-effective use of resources in view of other competing priorities such as expansion of antiretroviral therapy coverage. We assessed the cost-effectiveness of alternative patient monitoring strategies.

Methods We evaluated a range of monitoring strategies, including clinical, CD4 cell count, and viral load monitoring, alone and together, at different frequencies and with different criteria for switching to second-line therapies. We used three independently constructed and validated models simultaneously. We estimated costs on the basis of resource use projected in the models and associated unit costs; we quantified impact as disability-adjusted life years (DALYs) averted. We compared alternatives using incremental cost-effectiveness analysis.

Findings All models show that clinical monitoring delivers significant benefit compared with a hypothetical baseline scenario with no monitoring or switching. Regular CD4 cell count monitoring confers a benefit over clinical monitoring alone, at an incremental cost that makes it affordable in more settings than viral load monitoring, which is currently more expensive. Viral load monitoring without CD4 cell count every 6–12 months provides the greatest reductions in morbidity and mortality, but incurs a high cost per DALY averted, resulting in lost opportunities to generate health gains if implemented instead of increasing antiretroviral therapy coverage or expanding antiretroviral therapy eligibility.

Interpretation The priority for HIV programmes should be to expand antiretroviral therapy coverage, firstly at CD4 cell count lower than 350 cells per μL, and then at a CD4 cell count lower than 500 cells per μL, using lower-cost clinical or CD4 monitoring. At current costs, viral load monitoring should be considered only after high antiretroviral therapy coverage has been achieved. Point-of-care technologies and other factors reducing costs might make viral load monitoring more affordable in future.

Funding Bill & Melinda Gates Foundation, WHO.

Introduction

The monitoring of patients on antiretroviral therapy is an important part of HIV care: it determines whether treatment is successful, or if a different drug regimen or improved adherence is required. Patients with treatment failure are more likely to have progressive disease and are at greater risk of dying, and patients with non-suppressed virus are also at risk of developing resistance and transmitting HIV infections to others. Patients can be monitored and treatment failure can be defined in many ways, in terms of the assays used (clinical monitoring with or without the measurement of CD4 count or plasma viral load), the frequency of checks (eg, every 3, 6, 12, or 36 months), and the decision rules applied for change of antiretroviral therapy based on clinical, CD4 count, or viral load criteria.

Every monitoring strategy carries different costs and health consequences. Determination of the cost effectiveness of a given strategy requires decision makers to balance the gains in health it provides against the gains in health that could be achieved by allocating resources to other interventions. Health-economic analyses such as those presented here can provide guidance on how to measure and value health outcomes, and on how to allocate scarce resources to generate health gains at the population level.

Since 2006, WHO antiretroviral therapy guidelines have recommended a "public health approach" to antiretroviral therapy scale-up,[1,2] based on standardised and simplified treatment and monitoring. This approach includes a common first-line regimen of a non-nucleoside reverse transcriptase (NNRT) inhibitor plus two nucleoside reverse transcriptase (NRT) inhibitors, of which one should be either zidovudine or tenofovir, that can be delivered in decentralised settings. The 2010 guidelines recommended that patients receive regular

New South Wales, Sydney, NSW, Australia (R Gray PhD); University of Liverpool, Liverpool, UK (A Hill PhD); Center for Health Decision Science, Harvard School of Public Health, Boston, MA, USA (N A Menzies MPH); and HIV Programme, WHO, Geneva, Switzerland (P Easterbrook MB, M Doherty MD, G Hirnschall MD)

Correspondence to:
Prof Timothy B Hallett, School of Public Health, Imperial College London, Norfolk Place, St Mary's Campus, London W2 1PG, UK
timothy.hallett@imperial.ac.uk

See **Online** for appendix

clinical and immunological monitoring and, if feasible, virological monitoring and then switch to a different drug regimen (a second-line) once treatment failure is detected using any one of the following criteria: (1) clinical failure—ie, a new or recurrent WHO stage 4 event; (2) immunological failure—ie, a fall of CD4 count to baseline (the start of treatment) or below, a 50% fall from an on-treatment peak value, or persistent CD4 levels below 100 cells per µL; and (3) virological failure—ie, plasma viral load higher than 5000 copies per mL while on treatment. These three definitions can be detected with clinical monitoring, measurement of CD4 cell counts, and viral load monitoring, respectively.

The 2013 WHO Consolidated ARV Guidelines[3] no longer recommend the use of a 50% fall from on-treatment peak value for assessing immunological failure, and recommend viral load monitoring as the preferred approach to diagnose and confirm antiretroviral treatment failure in both adults and children, with failure defined as two consecutive plasma viral loads above 1000 copies per mL within three months of one another after six or more months on treatment. However, countries still need to decide whether programmes currently using clinical or CD4 monitoring should invest resources in upgrading clinics and laboratory infrastructure to use viral load monitoring.

Mathematical modelling and health economic evaluation allows for systematic, detailed consideration of the costs and health consequences of a broad repertoire of potential monitoring strategies over a range of timescales, and can therefore help inform countries' decisions on how to invest limited resources. Consequently, we collated evidence from published modelling studies and undertook new analyses on the cost effectiveness of alternative patient monitoring strategies. This study, which informed the WHO Guidelines, aims to identify appropriate monitoring strategies for programmes given their competing priorities and the wide variety of situations and resource constraints that they face.

Methods

Search criteria

To identify modelling groups for the WHO Guidelines process, we identified relevant modelling studies published between Sept 15, 2007, and Sept 15, 2012, through a search of PubMed/Medline and Google Scholar with search terms including "viral load monitoring", "patient monitoring", "cost-effectiveness", "mathematical modelling", "anti-retroviral therapy", "modelling patient monitoring", and "HIV treatment monitoring". A list of studies reviewed (including some before Sept 15, 2007) is included in the appendix. We included models in the review if they assessed the effect of patient monitoring strategies on health outcomes (treatment failure, viral loads, CD4 cell counts, clinical events or progression to AIDS) in a simulated population over time and also incorporated a cost-effectiveness analysis. We contacted groups with publications meeting these criteria in the previous 5 years to participate in the WHO Guidelines process.

Models

Given the importance and complexity of the question, it was important not to base findings on a single mathematical model but rather to assemble a set of independently constructed and validated models. We contacted six modelling groups; three agreed to undertake new analyses for the project,[4–8] whereas two[9,10] did not undertake new analyses but did contribute to the collective analysis presented here.

The mathematical models used were: the HIV Synthesis model (Phillips and colleagues,[4,5] University College London, London, UK), Estill and colleagues[6,7] (University of Bern, Bern, Switzerland), and Braithwaite and colleagues[8,11] (New York University, NY, USA). The Synthesis and Estill models are parameterised for a generic southern Africa setting, whereas the Braithwaite model is parameterised for an east Africa setting. We assumed that the clinical progression of HIV was similar in these populations. Table 1 summarises key features of the models. Our paper focuses on the implementation of these models for health-economic analysis; see the appendix for further details on the models themselves.

We applied country-specific unit costs to each of these models to generate analyses for three countries representative of higher-resource, mid-resource, and lower-resource settings within southern Africa: South Africa, Zambia, and Malawi, respectively (appendix).

Choice of alternative monitoring strategies

We evaluated four sets of monitoring strategies, including clinical monitoring alone, CD4 cell count alone, routine viral load monitoring alone, at various thresholds; and two strategies comprising combinations of monitoring approaches. We also used a hypothetical scenario of no monitoring and no switching as a baseline comparator to establish the incremental costs

	Time horizon of simulation	Model tracks patients' morbidity and mortality	Model tracks HIV transmission from patients to others	Modelled outcomes related to patients' adherence to anti-retroviral therapy	Models include acquired and trans-mitted resistance
HIV Synthesis	15 years	Yes	Yes	Yes	Yes
Braithwaite and colleagues	20 years	Yes	No	Yes	Yes
Estill and colleagues	5 years	Yes	Not full transmission model, but calculated expected transmissions based on viral loads	Incorporated in scenario analysis using failure rate as a proxy	No

Table 1: **Features of the selected models**

and effects of each monitoring strategy in affecting population health. These strategies indicate the spectrum of monitoring approaches used in high-income, middle-income, and low-income settings, as well as potential new strategies. Not every model evaluated all strategies (table 2).

Costs and outcomes

We estimated costs from a health-sector perspective, in which only costs falling on the health system are included, and any wider, societal costs or benefits are not included. We projected health-care resource use in the models (number of clinic visits, number and type of monitoring tests, first-line and second-line antiretroviral drugs prescribed, additional health-care use associated with disease progression) and applied associated unit costs representative of health-care delivery to estimate the total costs of strategies (all cost assumptions are listed in the appendix). Unit costs included personnel time, building costs, training, and facility management; we incorporated programmatic (above facility) costs on the basis of proportional mark-ups on unit costs of resource inputs.

We summarised the health effects of the alternative strategies in the form of disability-adjusted life-years (DALYs) averted, a measure that captures the extent to which the interventions reduce the premature death and ill-health caused by a disease, including, in the HIV Synthesis model, reductions in morbidity and mortality from the prevention of onward HIV transmission.[5,6] DALYs averted for all scenarios run by each model are presented in the appendix. In our scenarios, 1 life-year in perfect health receives a weight of 0, whereas 1 life-year lived with a WHO stage 3 or 4 event (developed on the basis of viral load monitoring, CD4 cell count, and failure or absence of antiretroviral therapy) receives a substantial weight (0·547), and 1 life-year lived with asymptomatic HIV (eg, on successful antiretroviral therapy) receives a moderate weight of 0·053, portraying the decrement in the quality of life from these conditions.[12,13] We estimated DALYs averted by applying DALY-weights to years lived and clinical events generated by the models, and not through estimating total burden of disease in the countries. Both costs and outcomes are discounted to 2012 present value in US dollars using a 3% discount rate.[14]

Economic analyses

The expected costs and health outcomes (DALYs averted) associated with each of the monitoring alternatives can be compared to inform which is likely to represent the best value from available resources.[14,15] We ranked the strategies by effectiveness, removing those less effective and more costly than an alternative (ie, subject to dominance) or a linear combination of alternatives (subject to extended dominance). We compared all remaining strategies using incremental cost-effectiveness ratios (ICERs), showing the additional cost per unit of health gain (DALY averted) from a strategy compared with the next most effective alternative. ICERs are represented graphically in the form of cost-effectiveness frontiers that connect those strategies that provide the greatest health returns at any given cost.

	Threshold for switching	Abbreviation	Frequency of monitoring	Monitoring strategy included in the model		
				HIV Synthesis	Braithwaite	Estill
No monitoring	None (no switching)	NS	None (no monitoring)	Implemented*	Implemented*	Implemented*
Clinical monitoring	WHO stage 4 event	CM, S4	Every 6 months	Implemented*	Implemented*	No
Clinical monitoring	WHO stage 3 or 4 event	CM, S3/4	Every 6 months	Implemented*	No	Implemented*
Clinical monitoring and CD4 cell count	CD4 <100 cells per μL or new stage 4 event	CD4 <100/S4	Every 6 months	Implemented*	No	No
CD4 cell count	CD4 cell count below baseline or <50% of peak value on ART	CD4-CA	Every 6 months	Implemented*	Implemented*	Implemented*
CD4 cell count and viral load monitoring	CD4 cell count below baseline or <50% of peak value on ART; VL ≥1000 copies per mL	CD4/TGVL	Every 6 months (CD4 count—VL only done if CD4 failure)	Implemented*	Implemented*	Implemented*
Clinical monitoring plus CD4 cell count plus viral load monitoring	New stage 4 event; or CD4 cell count below baseline or <50% of peak value on ART; or viral load ≥1000 copies per mL	CD4/TGVL+	Every 6 months (Clinical monitoring plus TGVL); every 12 months (routine viral load monitoring)	Implemented*	No	No
Viral load monitoring	1000 copies per mL	VL12	Every 12 months	Implemented*	Implemented*	Implemented*
Viral load monitoring	1000 copies per mL	VL36	Every 36 months	Implemented*	No	Implemented*
Viral load monitoring	500 copies per mL	VL6/VL ≥500	Every 6 months		No	No
Viral load monitoring	1000 copies per mL	VL6	Every 6 months	Implemented*	Implemented*	Implemented*
Viral load monitoring	5000 copies per mL	VL6/VL ≥5000	Every 6 months	Implemented*	No	No
Viral load monitoring	10 000 copies per mL	VL6/VL ≥10 000	Every 6 months	Implemented*	Implemented*	No

TGVL=targeted viral load. ART=antiretroviral therapy. CA=current algorithm. *Scenario was implemented in corresponding model.

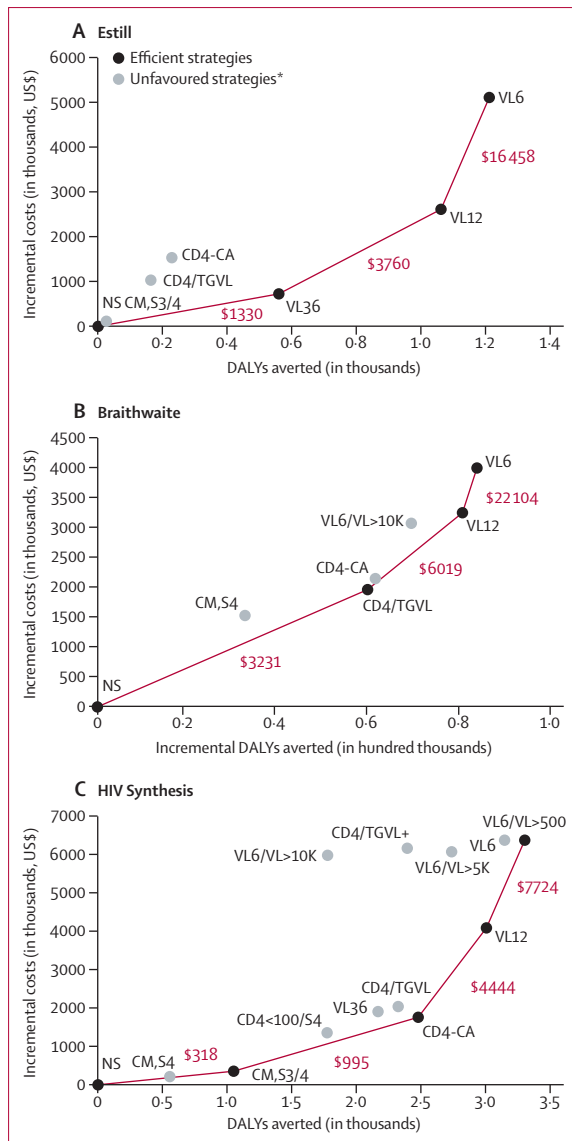*Table 2*: Monitoring strategies modelled, by abbreviation

Figure 1: Cost-effectiveness frontier plots for Zambia (ICERs per DALY averted, 2012 US$)

DALY=disability-adjusted life-year. (A) Estill model. (B) Braithwaite model. (C) HIV Synthesis model. The frontier line that represents a most efficient pathway of spending as resources increase is shown in red together with the ICERs—ie, the incremental cost per DALY averted of moving from one strategy to the next along the frontier. NS=no monitoring and no switching. VL6=viral load monitoring every 6 months. VL12=viral load monitoring every 12 months. VL36=viral load monitoring every 36 months. CM S3/4=clinical monitoring with switching on a new WHO stage 3 or 4 event. CD4-CA=clinical monitoring plus CD4 monitoring, switching at fall of CD4 cell count below baseline or of 50% or more from peak value on treatment. CD4/TGVL=targeted viral load strategy (viral load is used to confirm a suspected failure based on immunological criteria). >500=switching at >500 copies per mL. >5K=switching at >5000 copies per mL. >10K=switching at >10 000 copies per mL. CD4<100/S4=switching at <100 cells per μL or new stage 4 event. *Dominated or extendedly dominated strategies.

Importantly, ICERs alone cannot show which strategy is likely to be most appropriate for a particular setting: this requires comparison with a cost-effectiveness threshold.

The appropriate threshold in a particular setting depends on the opportunity costs of committing resources to fund an intervention, measured in terms of the health gains foregone because of displacement of alternative interventions that would not then be provided. An intervention can therefore only be deemed cost effective if the health gains that the intervention generates exceed what would have been gained if that intervention was not adopted and the resources were deployed elsewhere.

Opportunity costs themselves depend on the decision context and how else resources could be spent. In situations where scale-up of antiretroviral therapy is not complete, opportunity costs might include health gains from the provision of antiretroviral therapy using lower-cost monitoring approaches to those in need who are not currently receiving treatment. We therefore compared patient monitoring results with estimates of the cost effectiveness of antiretroviral therapy to infer the value-for-money of monitoring alternatives (ie, we compared the health benefits of money spent on monitoring with money spent on expanding antiretroviral therapy).

Sensitivity analyses investigate how results change with lower testing costs (as might be expected with the arrival of point-of-care or other new technologies) and reduced second-line antiretroviral drugs costs. These are presented in the form of incremental net monetary benefit (I-NMB), of routine viral load monitoring compared with the best monitoring alternative at a particular cost-effectiveness threshold. I-NMB is a measure of the value of health gains, on a monetarised scale, resulting from an intervention compared with the health gains that could be realised if the resources required to fund that intervention were used for alternative purposes.[15] A positive I-NMB for routine viral load monitoring therefore indicates it is cost effective compared with other monitoring alternatives at a given cost-effectiveness threshold, whereas a negative I-NMB indicates the health gains are not large enough relative to costs to recommend its adoption.

Role of the funding source
WHO authors contributed to the design of the study, the selection of settings considered and strategies evaluated, but had no role in the development or selection of epidemiological models, the conduct of the analyses, or interpretation of results. The Bill & Melinda Gates Foundation had no role in the design of the analysis, interpretation of the results, or the decision to submit the manuscript for publication. The corresponding author had final responsibility for the decision to submit for publication.

Results
The ICERs per DALY averted for each strategy are presented for Zambia in figure 1. Significantly, the results from Malawi and South Africa were in precise qualitative agreement in that the ranking of each
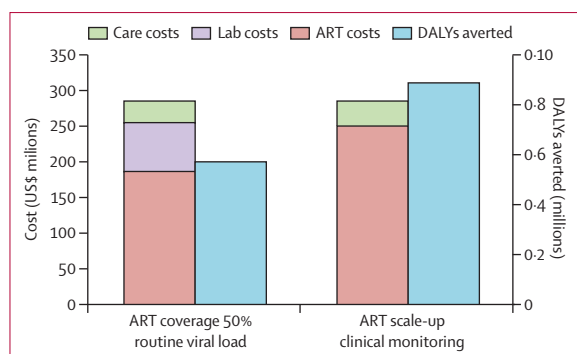
Figure 2: Costs and benefits (DALYS averted) of alternative uses of resources (Braithwaite model)
DALY=disability-adjusted life-year. ART=antiretroviral therapy. Results are per 100 000 HIV-infected individuals with both benefits and costs estimated over a 20 year budgeting horizon and discounted at 3% per annum.

Figure 3: Mean per-patient lifetime costs and DALYs-averted from alternative uses of ART treatment resources (Braithwaite model)
DALY=disability-adjusted life-year. ART=antiretroviral therapy. VL=viral load.

scenario along the cost frontier is the same across the three countries (appendix).

All models show that no monitoring and no switching (ie, maintaining one line of antiretroviral therapy) is the least costly and least effective strategy in the base case analyses. Viral load monitoring every 6 months (VL6) is the most costly and most effective alternative in every model; viral load monitoring every 12 months (VL12, switching at >1000 copies per mL) is the next-most-effective strategy in all models and is also slightly less costly.

Clinical and CD4-based monitoring approaches represent intermediate alternatives in cost and effective-ness in all models (figure 1). In the HIV Synthesis model, clinical monitoring (switching on a new WHO stage 3 or 4 event [CM S3/4]) offers notable benefits at low incre-mental costs compared with no monitoring and no switching. The addition of CD4 monitoring (CD4-CA) to clinical monitoring alone confers a benefit particularly in the HIV Synthesis and Braithwaite models, at an incremental cost meaning that it might be affordable in more settings than regular viral load monitoring. The Braithwaite results lend support to a targeted viral load strategy (CD4/TGVL, whereby a viral load is used to confirm a suspected failure based on immunological criteria), which might be considered as a stepping stone towards the routine use of viral load monitoring—perhaps as programmes wait for cheaper point-of-care viral load monitoring to become widely available. This strategy would be less likely to be favoured, however, if it meant new viral load laboratory infrastructure had to be built or if it led to viral load machines being used at low volume and higher unit costs. Furthermore, we note that CD4-CA lies very close to the frontier in the Braithwaite model.

To assess whether improvements in patient monitoring should be prioritised over expanded coverage of antiretroviral therapy, we ran the Braithwaite model using costs from Malawi. We assumed that the antiretroviral therapy coverage (ie, the proportion of people eligible for
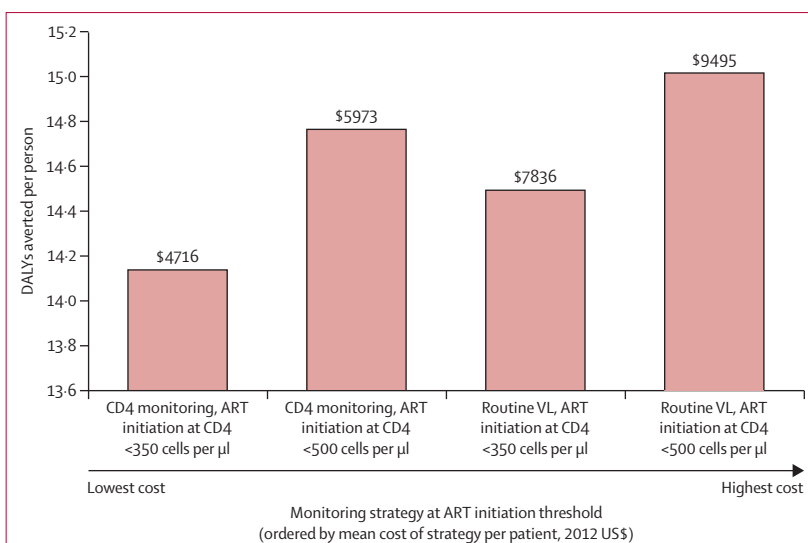
antiretroviral therapy who are receiving it) was currently 50%, and that clinical monitoring was used for patients on antiretroviral therapy. In these respects, the model represents the situation in many eastern and southern African countries, where despite recommendations for CD4 or viral load monitoring being in place, scale-up of these strategies is limited and clinical monitoring remains widespread (median antiretroviral therapy coverage level noted in east and southern Africa is 56%[16]).

We considered a situation in which an HIV/AIDS programme has a choice between investing additional resources in routine 6-monthly viral load monitoring (while maintaining antiretroviral therapy coverage at 50%), or in increasing antiretroviral therapy coverage from 50% while still using clinical monitoring. In this hypothetical example, increasing antiretroviral therapy coverage—rather than upgrading patient monitoring—would be expected to generate much greater health benefits (figure 2). This result is consistent with the enormous benefits of antiretroviral therapy for patients with CD4 cell count of about 350 cells per μL compared with not receiving antiretroviral therapy at all, and the relatively modest benefits associated with the more extensive patient monitoring strategies in all the models (figure 1).

Other studies have also estimated that health gains for introducing antiretroviral therapy with clinical monitoring compared with no antiretroviral therapy can be realised at much lower ICERs than we estimate for the introduction of CD4 and viral load monitoring (Braithwaite, 2011, estimates an ICER of $600 per quality-adjusted life-year [QALY] gained for two lines of antiretroviral therapy with clinical monitoring and no fixed assumptions on the number of regimens available versus no antiretroviral
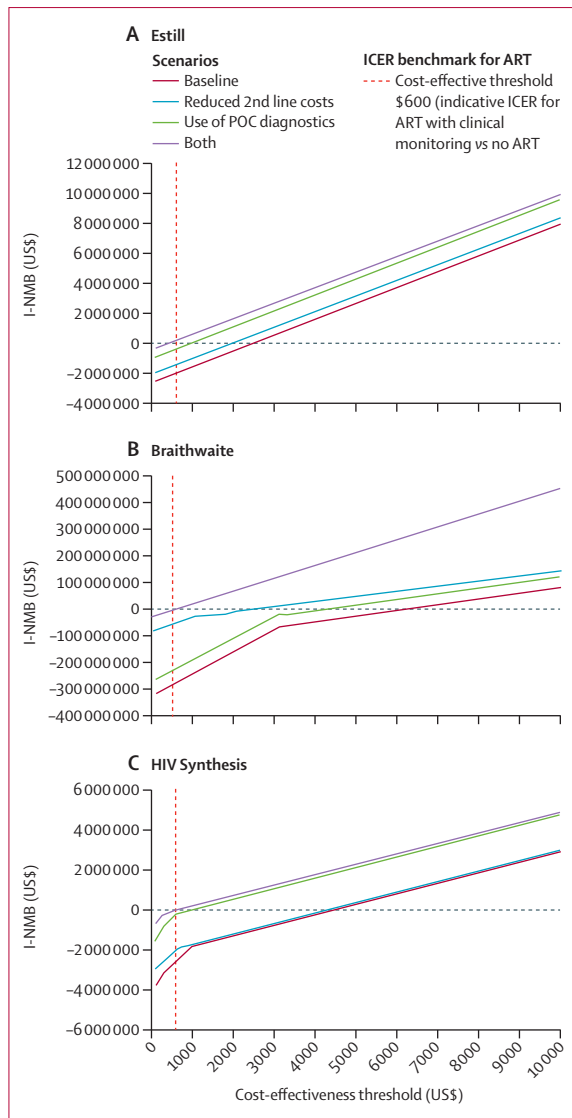
Figure 4: Scenario analyses

I-NMB=incremental net monetary benefit. POC=point of care. ICER=incremental cost-effectiveness ratio. ART=antiretroviral therapy. The figures show the I-NMB of 12-monthly routine viral load monitoring compared with the best alternative non-routine viral load strategy at a given cost-effectiveness threshold. A positive value of I-NMB implies that 12-monthly viral load monitoring (vertical axis) is cost effective at a particular cost-effectiveness threshold (horizontal axis), whereas a negative I-NMB indicates it is not cost effective because the opportunity costs exceed the health gains the intervention offers. Routine viral load monitoring becomes "cost effective" under each scenario at the threshold where the I-NMB line crosses the horizontal axis.

therapy, which is similar to previous published estimates of $590 per life-year gained,[17] and $628 per QALY gained with one line of antiretroviral therapy only[18]).

In some other settings with high antiretroviral therapy coverage using CD4 monitoring, such as Zambia, the relevant policy choice would seem to be whether to spend additional resources on the provision of viral load monitoring or increasing the antiretroviral therapy

eligibility criteria to CD4 cell count lower than 500 cells per µL. An additional analysis was run in the Braithwaite model to examine these alternatives (figure 3). The findings suggest that earlier initiation of antiretroviral therapy, while still using CD4 monitoring, would cost less and generate greater health gains than would keeping the threshold of antiretroviral therapy initiation at 350 cells per µL and using viral load monitoring. This finding is also indicated in the low ICERs (less than $290 per DALY averted in Zambia) that have recently been reported for earlier antiretroviral therapy initiation in those already in care.[19]

To assess the sensitivity of our results to particular cost assumptions, and to examine how results might change in response to changing costs, we constructed two alternative scenarios: (1) reduced costs of second-line antiretroviral regimens, with second-line costing the same as average current first-line antiretroviral regimen costs; and (2) reduced costs of providing assays, which might be expected with the development of new CD4 cell count and viral load technologies, including point of care tests, of $4 per test for CD4 cell count and $10 per test for viral load monitoring (compared with the $9·50 per test for CD4 cell count and $45 per test for viral load monitoring used in the original analyses shown in figure 1).

Figure 4 shows the I-NMB associated with routine 12-monthly viral load monitoring compared with the best non-routine viral load monitoring strategy.[14,15] The I-NMB of routine viral load monitoring can be interpreted as the difference in the value of health gains generated from routine viral load monitoring and the value of health gains foregone as a result of those resources required to fund this monitoring strategy being unavailable to deliver other interventions, at particular cost-effectiveness threshold levels. At higher cost-effectiveness thresholds, resources buy fewer health gains elsewhere in the health-care system and therefore the I-NMB of routine viral load monitoring increases. This might be the case, for instance, if a country has full antiretroviral therapy coverage and few other opportunities to generate health gains at low cost. However, at lower cost-effectiveness thresholds, the higher costs of routine viral load monitoring are of greater consequence because they displace investments in interventions that could offer health gains at low cost.

Reduced second-line costs and reduced testing costs would make 12-monthly viral load cost effective at lower cost-effectiveness thresholds than under base case assumptions (marked by where the lines cross the x-axis). However, the magnitude of these effects varies somewhat across the models. In the HIV Synthesis model, reduced second-line costs had very little effect on the cost effectiveness of routine viral load monitoring, but when the costs of the tests themselves fall, routine viral load monitoring becomes cost effective at a much lower threshold. In the Braithwaite and Estill models, reductions in the costs of second-line treatment were

more important for the cost effectiveness of viral load monitoring. However, in all models, the combination of reducing second-line cost and reduction in test costs makes it much more likely that viral load monitoring would be cost effective.

Discussion

This analysis shows that a limited availability of resources to monitor patients should not be a barrier to scale-up of antiretroviral therapy. We find that expanding treatment to more patients at existing thresholds for antiretroviral therapy initiation, or initiating antiretroviral therapy at higher CD4 cell counts while using clinical or immunological monitoring, would be a more effective use of resources than investing in more extensive patient monitoring using viral load tests. However, we also find that viral load monitoring would confer additional benefits to patients and populations, especially over the long term, and that if the cost of viral load monitoring falls substantially, then it might become a cost-effective strategy in future, particularly in settings with high antiretroviral therapy coverage (panel).

A major strength of this analysis is that it draws on various independent models, which come to very similar conclusions. This provides reassurance that the conclusions are robust to different ways in which the disease progression and monitoring can be represented in models. Although we have not provided results for the models across ranges of assumptions for adherence, delays in switching patients, and other factors, we know of no data that suggest these issues would interfere with the overall conclusions we have drawn about the relative priorities of the different strategies. We have not explicitly presented the impact of monitoring on HIV drug resistance or HIV transmission, but we emphasise that this is included in two of the models presented (table 1) and its effects are captured in the aggregate estimate of impact. Furthermore, it is possible that, by parameterising the models on the basis of trial data, the models overestimate the effect of monitoring compared with what would happen in real programmes. There is no reason to believe, however, that this would systematically bias our result to favouring one strategy over another, although there would be great benefit in evaluating the performance of these alternative strategies in routine programmes to test this assumption.

The systematic nature of our compiled analysis has afforded insights into the underlying reasons for the models to give slightly different results in some cases. Particularly, in the HIV Synthesis model, CD4 monitoring strategies perform better relative to other strategies than is the case in the two other models. This difference seems to be because, in that model, the proportion of life-years lived with immunological failure where there is also virological failure (at >1000 copies per mL) is higher than in the Braithwaite model (and higher than the proportion of concurrent

failures [as episodes, not life-years lived] in the Estill model). It is also higher than some reports in the literature of the positive predictive values of CD4 failure for virological failure, although these studies are based largely on people who initiated ART fairly recently, and the correlation between virological and immunological failure may increase with time on ART.[32–39] Thus, in the HIV Synthesis model, the CD4 information is assumed to be a more reliable guide to viral failure than elsewhere, reducing the marginal gains in health from using viral load monitoring in that model.

Other published models besides the three used here have also examined optimal strategies for antiretroviral therapy monitoring, in a range of settings, and have findings that are consistent with our results.[5,9,20] One model, however, stands in contrast: Hamers and colleagues[10] previously suggested that viral load monitoring would be cost-saving and could improve life-expectancy.

Panel: Research in context

Systematic Review

We searched PubMed, Medline, and Google Scholar for modelling studies published between Sept 15, 2007, and Sept 15, 2012, with search terms "viral load monitoring", "patient monitoring", "cost-effectiveness", "mathematical modelling", "antiretroviral therapy", "modelling patient monitoring", and "HIV treatment monitoring." A list of studies reviewed (including some before Sept 15, 2007) is shown in the appendix. Patient monitoring models were last reviewed in 2010.[20] Mathematical models that have attempted to represent disease progression and monitoring are consistent with trial and observational data: immunological monitoring offers some morbidity and mortality benefit (ie, less time spent with clinical events, fewer deaths) over clinical monitoring, and virological monitoring might offer some morbidity and mortality benefit over immunological monitoring.[4–8,11] Two randomised controlled trials have found that routine CD4 monitoring reduces patient morbidity and mortality relative to clinical monitoring alone.[21–23] Several studies have evaluated the added effect of viral load monitoring compared with CD4 or clinical monitoring, but have not found major effects on morbidity or mortality.[21,22,24–27] However, compared with CD4 monitoring[21,24] or clinical monitoring,[26,27] routine CD4 and viral load monitoring led to more patient switching to second-line drugs. Routine use of viral load was found to lead to more frequent switches to second-line drugs, compared with use of viral load only to confirm a failure based on clinical or immunological criteria.[25] It has also been suggested that viral load monitoring (and by implication, targeted viral load for confirmation of immunological failure) might prevent unnecessary switches to second-line therapy in patients who are failing clinically or immunologically but not virologically.[28] Less time spent with non-suppressed viral load could reduce the development of resistance[5,28,29] and the onward transmission of HIV;[5,30,31] however, Laurent and colleagues[26,27] found no difference in the proportion of resistance in the clinical and laboratory arms, and Jourdain and colleagues[23] found just one case with resistance mutations in the CD4 arm.

Interpretation

Drawing overarching conclusions from existing patient monitoring models is complicated by models' use of different cost inputs and heterogeneity in strategies modelled. Our analysis shows that three models brought together and run on a core set of scenarios with the same costs come to largely similar conclusions. It also confirms that these modelling results are largely consistent with the trial literature, and over a longer timeframe than the trial data.

Two reasons might explain this discrepancy. First, the model does not model clinical or immunological failure without virological failure; second, clinical and CD4 monitoring therefore underperform because they are assumed to have no intrinsic value beyond correlating (weakly, in this model) with viral failure.

Programmes need to decide how to use available resources for the benefit of the populations they serve. Unfortunately, not all health-care interventions that offer health gains can be funded, and adoption of interventions that require additional resources means that these resources are then unavailable for the delivery of other interventions that could also generate health gains. Decision makers therefore need to determine a value at which the costs per health gains associated with a more effective, but more expensive, intervention are deemed acceptable such that committing resources to that intervention is likely to improve population health. One threshold used by WHO is that any intervention that generates a unit of health gain (DALY averted) at less than three times gross domestic product (GDP) per capita of a country is "relatively cost effective" and anything less than GDP per capita "highly cost effective".[40] However, there is little evidence to support these thresholds and so instead we compare results to the health gains that could be achieved through committing resources to the expansion of antiretroviral therapy coverage. There might, however, be other opportunity costs both within the realm of HIV/AIDS services (such as using different drug regimens) as well as in other areas of the health system or even in other sectors entirely.

These analyses are intended to contribute to deliberative processes of resource prioritisation. There are also likely to be other policy goals in addition to maximising health gains. A concern for equity, for example, could favour the adoption of a cheaper but less-effective monitoring strategy if its lower cost means that monitoring can be delivered to a greater number of people.[41] Specific practical considerations will also be important, such as existing laboratory infrastructure capabilities and the timing of procurement cycles. Furthermore, these decisions should be re-evaluated as antiretroviral therapy programmes expand and new diagnostics are developed or prices are reduced. The anticipated future availability of point-of-care technologies is likely to be particularly important, since those new tests might provide even greater benefits than our analysis of their potential lower cost implies, such as allowing more rapid generation and delivery of results in more remote communities. Also, in settings where coverage of antiretroviral therapy at high CD4 cell counts increases, viral load monitoring might enhance the impact that antiretroviral therapy has on reducing HIV transmission while the usefulness of current CD4 monitoring algorithms might be reduced.

The key question for programmes is not whether viral load monitoring provides benefit to patients. Rather, the question is whether, given available programme resources, the relatively modest anticipated benefit of viral load monitoring is worth the added cost, and whether the opportunity costs in morbidity and mortality of forgoing the use of these resources for other efforts is acceptable. We show here that routine viral load monitoring at current cost might be appropriate only in wealthier countries, especially those that have scaled-up to close-to-full antiretroviral therapy coverage, or if the cost of viral load testing were to fall considerably.

References
1   Gilks CF, Crowley S, Ekpini R, et al. The WHO public health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet* 2006; **368:** 505–10.
2   WHO. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach—2010 revision. Geneva: World Health Organization, 2010.
3   WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Geneva: World Health Organization, 2013.
4   Phillips AN, Pillay D, Miners AH, Bennett DE, Gilks CF, Lundgren JD. Outcomes from monitoring of patients on antiretroviral therapy in resource-limited settings with viral load, CD4 cell count, or clinical observation alone: a computer simulation model. *Lancet* 2008; **371:** 1443–51.
5   Phillips AN, Pillay D, Garnett G, et al. Effect on transmission of HIV-1 resistance of timing of implementation of viral load monitoring to determine switches from first to second-line regimens in resource-limited settings. *AIDS* 2011; **25:** 843–50.
6   Estill J, Aubrière C, Egger M, et al. 2012. Viral load monitoring of antiretroviral therapy, cohort viral load and HIV transmission in Southern Africa: a mathematical modelling analysis. *AIDS* 2012; **26:** 1403–13.
7   Estill J, Egger M, Johnson LF, et al. Monitoring of antiretroviral therapy programmes in Malawi, South Africa and Zambia: Mathematical Modelling Study. *PLoS One* 2013; **8:** e57611.15.
8   Braithwaite RS, Nucifora KA, Yiannoutsous CT, et al. Alternative antiretroviral monitoring strategies for HIV-infected patients in east Africa: opportunities to save more lives? *J Intl AIDS Soc* 2011; **14:** 38.

9 Kimmel AD, Weinstein MC, Anglaret X, et al. Laboratory monitoring to guide switching antiretroviral therapy in resource-limited settings: clinical benefits and cost-effectiveness. *J Acquir Immune Defic Syndr* 2010; **54:** 258–68.

10 Hamers RL, Sawyer AW, Tuohy M, Stevens WS, Rinke de Wit TF, Hill AM. Cost-effectiveness of laboratory monitoring for management of HIV treatment in sub-Saharan Africa: a model-based analysis. *AIDS* 2012; **26:** 1663–72.

11 Braithwaite RS, Nucifora KA, Toohey C, et al. How do different eligibility guidelines for antiretroviral therapy affect the cost-effectiveness of routine viral load testing in sub-Saharan Africa? Submitted; AIDS (2013). In press.

12 Murray CJ, Acharya AK. Understanding DALYs (disability-adjusted life years). *J Health Econ* 1997; **16:** 703–30.

13 Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* 2012; **380:** 2129–4310.

14 Acharya A, Adam T, Baltussen R. Making choices in health: WHO guide to cost effectiveness analysis. Geneva: World Health Organization, 2004.

15 Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, USA, 2005.

16 Joint United Nations Programme on HIV/AIDS. UNAIDS Data Tables 2011 [Internet]. Geneva: World Health Organization, 2011. http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/JC2225_UNAIDS_datatables_en.pdf (accessed May 15, 2013).

17 Goldie SJ, Yazdanpanah Y, Losina E, et al. Cost-effectiveness of HIV treatment in resource-poor settings—the case of Côte d'Ivoire. *N Engl J Med* 2006; **355:** 1141–53.

18 Bishai D, Colchero A, Durack DT. The cost effectiveness of antiretroviral treatment strategies in resource-limited settings. *AIDS* 2007; **21:** 1333–40.

19 Eaton JW, Menzies NA, Stover J, et al. Health benefits, costs, and cost-effectiveness of earlier eligibility for adult antiretroviral therapy and expanded treatment coverage: a combined analysis of 12 mathematical models. *Lancet Glob Health* 2014; **2:** e23–34.

20 Walensky RP, Ciaranello AL, Park JE, Freedberg KA. Cost-effectiveness of laboratory monitoring in sub-Saharan Africa: a review of the current literature. *Clin Infect Dis* 2010; **51:** 85–92.

21 Mermin J, Ekwaru JP, Were W, et al. Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *BMJ* 2011; **343:** d6792.

22 Kahn JG, Marseille E, Moore D, et al. CD4 cell count and viral load monitoring in patients undergoing antiretroviral therapy in Uganda: cost effectivness study. *BMJ* 2011; **343:** d6884.

23 DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* 2010; **375:** 123–31.

24 Jourdain G, Ngo-Giang-Huong N, Le Coeur S, et al. PHPT-3: A randomized clinical trial comparing CD4 vs viral load ART monitoring/switching strategies in Thailand. 18th Conference on Retroviruses and Opportunistic Infections; Boston, MA, USA; Feb 27–March 3, 2011. Abstr 44.

25 Saag A, Westfall A, Luhanga D, et al. Cluster randomized trial of routine vs discretionary viral load monitoring among adults starting ART: Zambia. 19th Conference on Retroviruses and Opportunistic Infections; Seattle, WA, USA; March 5–8, 2012.

26 Laurent C, Kouanfack C, Laborde-Balen G, et al. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. *Lancet Infect Dis* 2011; **11:** 825–33.

27 Boyer S, March L, Kouanfack C, et al. Monitoring of HIV viral load, CD4 cell count, and clinical assessment versus clinical monitoring alone for antiretroviral therapy in low-resource settings (Stratall ANRS 12110/ESTHER): a cost-effectiveness analysis. *Lancet Infect Dis* 2013; **13:** 577–86.

28 Sigaloff KCE, Hamers RL, Wallis CL, et al. Unnecessary antiretroviral treatment switches and accumulation of HIV resistance mutations; two arguments for viral load monitoring in Africa. *J Acquir Immune Defic Syndr* 2011; **58:** 23–31.

29 Reynolds SJ, Sendagire H, Newell K, et al. Virologic versus immunologic monitoring and the rate of accumulated genotypic resistance to first-line antiretroviral drugs in Uganda. *BMC Infect Dis* 2012; **12:** 381.

30 Attia S, Egger M, Muller M, Zwahlen M, Low N. Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis. *AIDS* 2009; **23:** 1397–404.

31 Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011; **365:** 493–505.

32 Badri M, Lawn SD, Wood R. Utility of CD4 cell counts for early prediction of virological failure during antiretroviral therapy in a resource-limited setting. *BMC Infect Dis* 2008; **8:** 89.

33 Chariwarith R, Wachirakaphan C, Kotarathititum W, et al. Sensitivity and specificity of using CD4+ measurement and clinical evaluation to determine antiretroviral treatment failure in Thailand. *Intl J Infect Dis* 2007; **11:** 413–16.

34 Keiser O, McPhail P, Boulle A, et al. Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy. *Trop Med Int Health* 2009; **14:** 1220–25.

35 Kiragga AN, Castelnuovo B, Kamya MR, Moore R, Manabe YC. Regional differences in predictive accuracy of WHO immunologic failure criteria. *AIDS* 2012; **26:** 768–70.

36 Mee P, Fielding KL, Charalambous S, Churchyard GJ, Grant AD. Evaluation of the WHO criteria for antiretroviral treatment failure among adults in South Africa. *AIDS* 2008; **22:** 1971–77.

37 Moore DM, Awor A, Downing R, et al. CD4+ T-cell count monitoring does not accurately identify HIV-infected adults with virologic failure receiving antiretroviral therapy. *J Acquir Immune Defic Syndr* 2008; **49:** 477–84.

38 Rawizza HE, Chaplin B, Meloni ST, et al. Immunologic criteria are poor predictors of virologic outcome: implications for HIV treatment monitoring in resource-limited settings. *Clin Infect Dis* 2011; **53:** 1283–90.

39 Reynolds SJ, Nakigozi G, Newell K, et al. Failure of immunologic criteria to appropriately identify antiretroviral treatment failure in Uganda. *AIDS* 2009; **23:** 697–700.

40 Choosing interventions that are cost-effective: cost-effectiveness thresholds. Geneva: World Health Organization, 2012. http://www.who.int/choice/costs/CER_thresholds/en/index.html (accessed April 30, 2012).

41 Walensky RP, Wood R, Ciaranello AL, et al. Scaling up the 2010 World Health Organization HIV treatment guidelines in resource-limited settings: a model-based analysis. *PLoS Med* 2010; **7:** e1000382.