How Should HIV Programmes Monitor Adults on ART? A Combined Analysis of Three Mathematical Models

Daniel Keebler[1^] & Paul Revill[2^], Scott Braithwaite[3], Andrew Phillips[4], Nello Blaser[5], Annick Borquez[6], Valentina Cambiano[4]*, Andrea Ciaranello[7]*, Janne Estill[5]*, Richard Gray[8]*, Andrew Hill[9]*, Olivia Keiser[5]*, Jason Kessler[3]*, Nicolas A Menzies[10]*, Kimberly A Nucifora[3]*, Luisa Salazar Vizcaya[5]*, Simon Walker[2]*, Alex Welte[1]*, Philippa Easterbrook[11], Meg Doherty[11], Gottfried Hirnschall[11] & Timothy B Hallett[6+]

^Contributed equally
*Listed in alphabetical order
[+] Corresponding author

[1] South African Department of Science and Technology/National Research Foundation Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa
[2] Centre for Health Economics, University of York, York, United Kingdom
[3] School of Medicine, New York University, New York, United States
[4] Research Department of Infection and Population Health, University College London, London, United Kingdom
[5] Division of International and Environmental Health, Institute of Social and Preventive Medicine (ISPM) University of Bern, Bern, Switzerland
[6] Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom
[7] Massachusetts General Hospital, Boston, Massachusetts, United States
[8] The Kirby Institute, The University of New South Wales, Sydney NSW 2052, Australia
[9] University of Liverpool, Liverpool, United Kingdom
[10] Center for Health Decision Science, Harvard School of Public Health, Boston, Massachusetts, United States
[11] HIV Programme, World Health Organization, Geneva, Switzerland

[+]Corresponding Author:
School of Public Health
Imperial College London
Norfolk Place
St Mary's Campus
London W21PG
United Kingdom
Tel: +44(0)20 7594 1150
E-mail: timothy.hallett@imperial.ac.uk

ABSTRACT

***Background*:** The WHO's 2013 revisions to its Consolidated Guidelines on ARVs will recommend routine viral load monitoring (VLM), 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 given other competing priorities, including expansion of ART coverage. Here we assess the cost-effectiveness of alternative patient monitoring strategies.

Methods: A range of monitoring strategies was evaluated, including clinical, CD4 and viral load monitoring alone and together at different frequencies and with different criteria for switching to second-line therapies. Three independently-constructed and validated models were analysed simultaneously. Costs were estimated based on resource use projected in the models and associated unit costs; impact was quantified as disability-adjusted life years (DALYs) averted. Alternatives were compared using incremental cost-effectiveness analysis.

Results: All models show that clinical monitoring delivers significant benefit compared to no monitoring and 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 VLM, which is currently more expensive. VLM without CD4 every six to 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 ART coverage or expanding ART eligibility.

Interpretation: The priority for HIV programmes should be to expand ART coverage, firstly at CD4 <350 cells and then at CD4 <500, using lower-cost clinical or CD4 monitoring. At current costs, VLM should be considered only after high ART coverage has been achieved.

Point-of-care technologies and other factors reducing costs may make VLM more affordable in future.

Funding: The HIV Modelling Consortium is funded by the Bill and Melinda Gates Foundation. Funding for this work was also provided by the World Health Organization.

Abstract: 300 words

Manuscript: 3962 words (excluding text in tables, figures and panels)

INTRODUCTION

Monitoring patients receiving ART 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 experience progressive disease and are at greater risk of dying, while patients with non-suppressed virus are also at risk of developing resistance and transmitting HIV infections to others. There are many different ways in which patients can be monitored and in which treatment failure can be defined, in terms of the assays used (clinical monitoring with or without the measurement of CD4 cell count and/or plasma viral load), the frequency of checks (e.g. every 3, 6, 12 or 36 months) and the decision rules applied for switching of ART based on clinical, CD4 or viral load criteria.

Each monitoring strategy carries different costs and health consequences (see Research in Context panel in the Discussion). Determining 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 ART guidelines have recommended a "public health approach" to ART scale-up[1,2], based upon standardized and simplified treatment and monitoring. This includes a common first-line regimen of a non-nucleoside reverse transcriptase inhibitor plus two nucleoside reverse transcriptase inhibitors, of which one should be either zidovudine (AZT) or tenofovir (TDF) that can be delivered in decentralized settings. Guidelines have recommended that patients receive regular 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:

- Clinical failure: a new or recurrent WHO Stage 4 event
- Immunologic failure: Fall of CD4 to baseline (i.e. CD4 count at start of treatment) or below; or 50% fall from an on-treatment peak value or persistent CD4 levels below 100
- Virologic failure: Plasma viral load (VL) over 5000 copies/mL while on treatment

These three definitions can be detected with clinical monitoring, measurement of CD4 cell counts, and viral load monitoring (VLM), respectively.

The 2013 World Health Organization (WHO) Consolidated ARV Guidelines will recommend viral load monitoring as the preferred monitoring approach to diagnose and confirm ARV treatment failure in both adults and children. However, it remains for countries 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 and detailed consideration of the costs and benefits of a broad potential repertoire of 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 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

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. Our search and selection criteria are outlined in the "Research in Context" panel in the Discussion section. Six modelling groups were contacted and three agreed to undertake new analyses for the project, while two (CEPAC and Hamers et al.) did not undertake new analyses but did contribute to the collective analysis presented here.

The mathematical models used were: HIV Synthesis model (Phillips, Cambiano et al., University College London)[3], Estill et al., (University of Bern)[4,5] and Braithwaite et al. (New York University)[6,7]. The Estill and Phillips models are parameterized for a generic "Southern Africa" setting, while Braithwaite is parameterized for an "East Africa" setting. It was assumed that the clinical progression of HIV is similar in these populations. Key features of the models are summarized in Table 1. Our paper focuses on the implementation of these models for health-economic analysis; see Appendix A.1 and Table A1 for further detail on the models themselves.

Model	Time horizon of simulation	Model tracks patient morbidity and mortality	Model tracks HIV transmission from patients to others	Modelled outcomes related to patients adherence to ART	Models include acquired and transmitted resistance
HIV Synthesis	15 years	Yes	Yes	Yes	Yes
Braithwaite	20 years	Yes	No	Yes	Yes

Estill	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

Country-specific unit costs were applied to each of these models to generate analyses for three countries representative of higher-, mid- and lower-resource settings within Southern Africa: South Africa, Zambia and Malawi (see Appendix A, Table A3, for full results for all three countries).

Choice of Alternative Monitoring Strategies

We evaluated four sets of monitoring strategies, including clinical monitoring alone, CD4 alone, routine viral load alone at various thresholds, and two strategies comprising combinations of monitoring approaches. A hypothetical no monitoring or 'no switching' scenario was also used as a baseline comparator to establish the incremental costs and effects of each monitoring strategy in affecting population health. These strategies reflect the spectrum of monitoring approaches employed in high, middle and low-income settings, as well as potential new strategies. Not every model evaluated all strategies (Table 2).

Abbreviation	**Diagnostic**	**Threshold for switching**	**Frequency of monitoring**	***Monitoring strategy included in the model***		
				HIV Synthesis	*Braithwaite*	*Estill*
NS	None (no monitoring)	None (no switching)	None (no monitoring)	X	X	X
CM, S4	Clinical	WHO Stage 4 event	6 months	X	X	
CM, S3/4	Clinical	Stage 3 or 4 event	6 months	X		X
CD4 <100/S4	Clinical + CD4	CD4 <100 or new stage 4 event	6 months	X		

CD4-CA*	CD4	Below baseline or <50% of peak value on ART	6 months	X	X	X
CD4/TGVL**	CD4 + VL	CD4 below baseline or <50% of peak value on ART; VL ≥ 1000 copies/mL	6 months	X	X	X
CD4/TGVL+	Clinical + CD4 + VL	New stage 4 event; or CD4 below baseline or <50% of peak value on ART; or VL ≥ 1000 copies/mL	6 months (Clinical + TGVL) 12 months (routine VL)	X		
VL12	VL	1000 copies/mL	12 months	X	X	X
VL36	VL	1000	36 months	X		X
VL6/VL≥500	VL	500	6 months			
VL6	VL	1000	6 months	X	X	X
VL6/VL≥5000	VL	5000	6 months	X		
VL6/VL≥10000	VL	10000	6 months	X	X	

* CA = Current Algorithm
** TGVL = Targeted Viral Load
X = Scenario was implemented in corresponding model.

Table 2: Monitoring Strategies Modelled

Costs and Outcomes

Costs were estimated from a health-sector perspective, where only costs falling on the health system are included and any wider societal benefits are not included. Health care resource use was projected in the models (number of clinic visits, number and type of monitoring tests, 1st and 2nd line ARVs prescribed, additional health care use associated with disease progression) and associated unit costs representative of healthcare delivery were applied to estimate the total costs of strategies (all cost assumptions are listed in Appendix A, Table A2). Unit costs incorporated personnel time, building costs, training and facility

management; programmatic ('above facility') costs were incorporated based on proportional mark-ups on unit costs of resource inputs.

The health impact of the alternative strategies was summarized in the form of disability-adjusted life years (DALYs) averted, a measure which captures the extent to which the interventions reduce the premature death and ill-health caused by the disease, including in the Synthesis model where the intervention reduces morbidity/mortality by preventing onward HIV transmission[8]. DALYs averted for all scenarios in each model are presented in the Appendix, Table A3. In our scenarios, one life-year in perfect health receives a weight of 0, while one life-year lived with a WHO Stage 3 or 4 event (developed based on viral load, CD4 and failure or lack of ART) receives a substantial weight (0.547), and one life-year lived with asymptomatic HIV (e.g. on successful ART) receives a moderately weight of 0.053, reflecting the decrement in the quality of life from these conditions[9]. Both costs and outcomes are discounted to 2012 present value in US dollars using a 3% discount rate[10].

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[11]. The strategies were ranked by effectiveness, removing those less effective and more costly than an alternative (i.e. subject to 'dominance') or a linear combination of alternatives (subject to 'extended dominance'). All remaining strategies were compared using incremental cost-effectiveness ratios (ICERs), showing the additional cost per unit of health gain (DALY-averted) from a strategy compared to the next most effective alternative. ICERs are shown graphically in the form of cost-effectiveness frontiers that connect those strategies which provided the greatest health returns for the least cost.

Importantly, ICERs alone cannot show which strategy is likely to be most appropriate for a particular setting: this requires comparison to a cost-effectiveness threshold. The appropriate threshold in a particular setting depends upon the opportunity costs of committing resources to fund an intervention, measured in terms of the health gains foregone due to displacement of alternative interventions that go unfunded. 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.

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

Sensitivity analyses investigate how results change with lower testing costs (as may be expected with the arrival of point-of-care or other new technologies) and reduced second-line ARV costs. These are presented in the form of incremental net monetary benefit (I-NMB), of routine VLM compared to the best monitoring alternative at a particular cost-effectiveness threshold. NMB is a measure of the value of health gains, on a monetarized scale, resulting from an intervention compared to the health gains that could be realized if the resources required to fund that intervention were used for alternative purposes. A positive I-NMB for routine VLM therefore indicates it represents value for money relative to 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.

RESULTS

Base Case 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 scenario along the cost frontier is the same across the three countries (see Appendix A, Table A3).

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

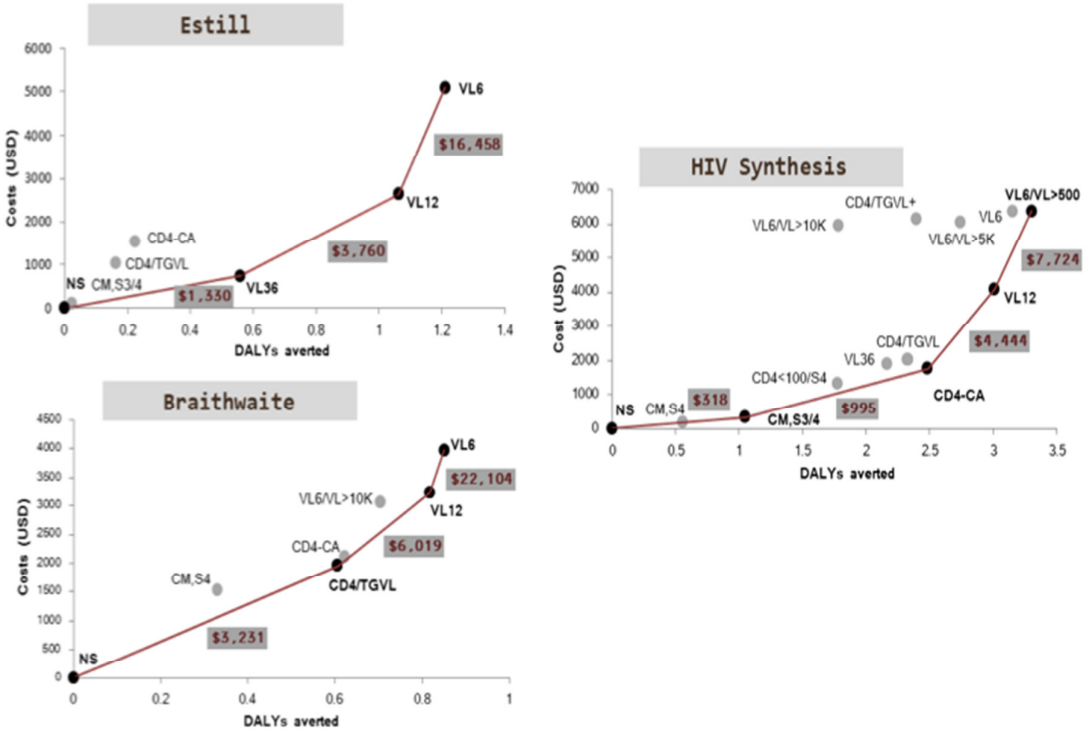

Figure 1: Cost-Effectiveness Frontier Plots for Zambia (ICERs per DALY Averted, 2012 US$)

(**a**) Estill model; (**b**) Braithwaite et al. model; (**c**) HIV Synthesis model. Unfavoured (i.e. dominated/extendedly dominated; see Methods) strategies are shown in light grey while most efficient strategies are shown in black and their code is highlighted in bold. The frontier line that represents a most efficient pathway of spending as resources increase is shown in red together with the ICERs, i.e. the incremental cost per DALY averted of moving from one strategy to the next along the frontier.

Clinical and CD4-based monitoring approaches represent intermediate alternatives in cost and effectiveness in all models (Figure 1). In HIV Synthesis, clinical monitoring (switching on a new WHO stage 3 or 4 event: CM S3/4) offers notable benefits at low incremental costs compared to no switching. The addition of CD4 monitoring (CD4-CA) to clinical monitoring alone confers a benefit particularly in HIV Synthesis and the Braithwaite model, at an incremental cost meaning that it may be affordable in more settings than regular viral load. 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 immunologic criteria), which may be considered as a "stepping stone" towards the routine use of viral load monitoring—perhaps as programs wait for cheaper point-of-care viral load to become widely available. This strategy would be less likely to be favoured; however, 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.

Opportunity Costs of Resource Allocations

In order to assess whether improvements in patient monitoring should be prioritized over expanded coverage of ART, the Braithwaite model was run using costs from Malawi (Figure 2). It was assumed that the ART coverage (i.e. the percentage of persons eligible for ART who are receiving it) was currently 50% and that clinical monitoring was used for patients on ART. 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 coverage level observed in East and Southern Africa is 56%[12]).

We considered a situation in which an HIV/AIDS programme has a choice between investing additional resources in routine 6-monthly viral load monitoring (whilst maintaining ART coverage at 50%), or in increasing ART coverage from 50% whilst still using clinical monitoring. In this hypothetical example, increasing ART 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 ART for patients with CD4 ~350 compared to not receiving ART 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 ART with clinical monitoring compared to no ART can be realized at much lower ICERs than we estimate for the introduction of CD4 and viral load monitoring (Braithwaite, 2011, estimates an ICER of $600/QALY for two lines of ART with clinical monitoring and no fixed assumptions on the number of regimens available versus no ART, which is similar to prior published estimates of $590/life year gained[13], and $628/QALY with one line of ART only[14]).

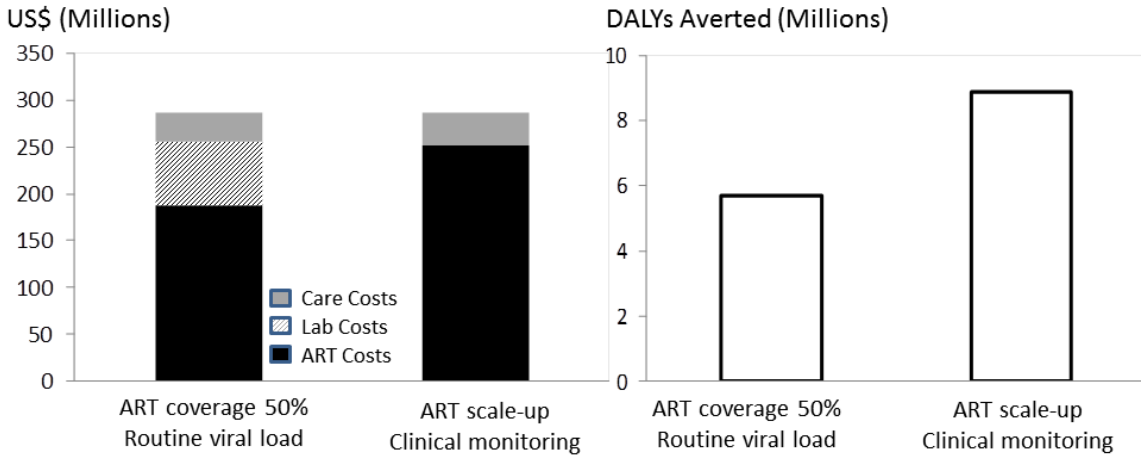

Figure 2: Costs and benefits (DALYS averted) of alternative uses of resources (Braithwaite model). Results given are per 1 million HIV-infected persons with both benefits and costs discounted at 3%.

In some other settings with high ART coverage using CD4 monitoring, such as Zambia, the relevant policy choice would appear to be whether to spend additional resources on providing viral load monitoring or increasing the ART eligibility criteria to CD4 <500. An additional analysis was run in the Braithwaite model to examine these alternatives (Figure 3). This suggests that earlier initiation of ART, whilst still using CD4 monitoring, would cost less and generate greater health gains than keeping the threshold of ART initiation at 350 and using viral load monitoring. This finding is also indicated in the low ICERs (less than $290/DALY averted in Zambia) that have recently been reported for earlier ART initiation among those already in care[15].

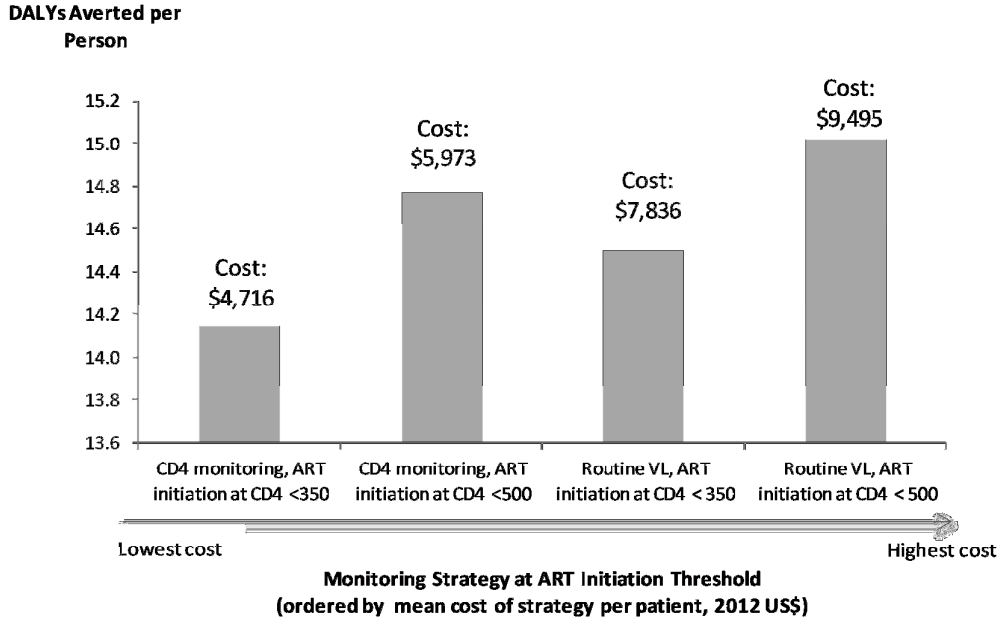

Figure 3: Costs per patient lifetime and DALYs-averted from alternative uses of resources (Braithwaite model).

Scenario Analyses

To assess the sensitivity of our results to particular cost assumptions, and to examine how results may change in response to changing costs, we constructed two alternative scenarios: (1) reduced costs of second-line ARV regimens, with second-line costing the same as average current first line ARV regimen costs; and (2) reduced assay costs, which may be expected with the development of new CD4 and VL technologies, including point of care tests, of $4 per test for CD4 and $10 per test for viral load (as compared to the $9.50 and $45 respectively used in the original analyses shown in Figure 1).

Figure 4 shows the incremental net monetary benefit (I-NMB) associated with routine 12-monthly viral load monitoring compared to all other non-viral load monitoring strategies[10,11]. The I-NMB of routine VLM can be interpreted as the difference in the value of health gains generated from routine VLM 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 VLM increases. This may be the case, for instance, if a country has full ART coverage and few other opportunities to generate health gains at low cost. However, at lower cost-effectiveness thresholds, the higher costs of routine VLM 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 a lower cost-effectiveness thresholds than under base case assumptions (marked by where the lines cross the x-axis). However, the magnitude of these impacts varies somewhat across the models. In the HIV Synthesis model, reduced second-line costs had very little impact 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 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.

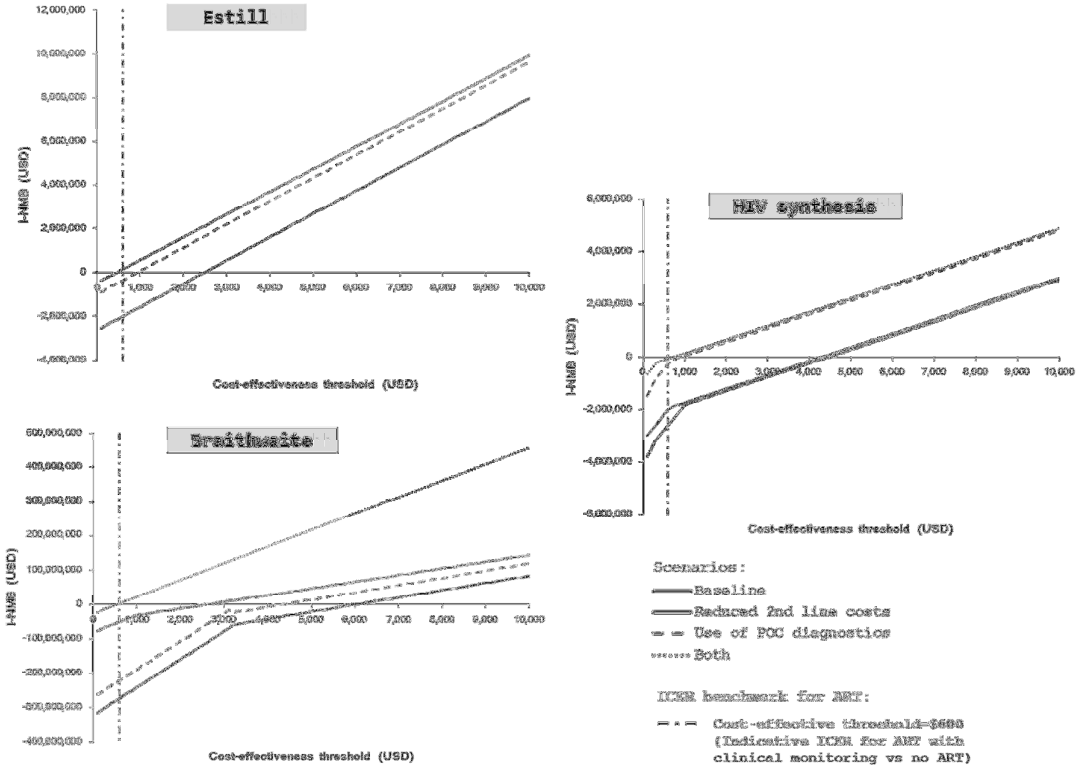

Figure 4: Scenario Analyses. The figures show the incremental net monetary benefit (I-NMB) of 12-monthly routine VL monitoring compared to the best alternative non-routine VL strategy at a given cost-effectiveness threshold. A positive value of I-NMB implies that 12-monthly viral load monitoring (vertical axis) is cost-effective at 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 VL monitoring becomes "cost-effective" under each scenario at the threshold where the I-NMB line crosses the horizontal axis.

DISCUSSION

This analysis shows that a limited availability of resources to monitor patients should not be a barrier to scale-up of ART. We find that expanding treatment to more patients at existing thresholds for ART initiation, or initiating ART 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 considerably, then it may become a cost-effective strategy in future, particularly in settings with high ART coverage.

Research in Context

Systematic Review

Patient monitoring models were last reviewed in 2010 by Walensky et al.[16] Mathematical models that have attempted to represent disease progression and monitoring are consistent with trial and observational data: immunologic monitoring offers some morbidity and mortality benefit (i.e. less time spent with clinical events, fewer deaths) over clinical monitoring, and virologic monitoring may offers some smaller morbidity and mortality benefit over immunologic monitoring[3,4,5,6,7,16]. Two randomized controlled trials have found that routine CD4 monitoring reduces patient morbidity and mortality relative to clinical monitoring alone[17,18]. Several studies have evaluated the added impact of viral load monitoring compared to CD4 or clinical monitoring, but not have found major impacts on morbidity or mortality[17,19,20,21]. However, compared with CD4 monitoring[17,19] or clinical monitoring[21], routine CD4 and viral load monitoring led to more patient switching to second-lines. Routine use of viral load was found to lead to more frequent switches to second-line, compared with using viral load only to confirm a failure based on clinical or immunological criteria[20]. It has also been suggested that viral load monitoring may prevent unnecessary switches to second-line therapy in patients who are failing clinically or immunologically but not virologically[22]. Less time spent with non-suppressed viral load could lead to less development of resistance[22,23] and reduced onward transmission of HIV[24,25]; however, Laurent et al. (21) found no difference in the proportion of resistance in the clinical and laboratory arms, and Jourdain et al. (19) found just one case with resistance mutations in the CD4 arm.

Search criteria

Relevant modelling studies published between 15 September 2007 and 15 September 2012 were identified through a search of PubMed/MEDLINE and Google Scholar, with search terms including "viral load monitoring," "patient monitoring," "cost-effectiveness patient monitoring," "modelling patient monitoring" and HIV "treatment monitoring." A list of studies reviewed (including some prior to 2008) is included in Appendix A, Table A1. Only cohort or population models of HIV treatment incorporating patient monitoring strategies as well as cost-effectiveness analyses were included. We invited all active modelling groups we found whose publications met these criteria to participate in the WHO Guidelines process.

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.

A major strength of this analysis is that it draws on a number of 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 emphasize 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 parametering the models based on trial data, the models over-estimate the impact of monitoring compared to what would happen in real programmmes. There is no reason to believe, however, that this would systematically bias our result to favoring one strategy over an other, 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 valuable insights into the underlying reasons for the models to give slightly different results in some cases. In particular, in the Synthesis model, CD4 monitoring strategies perform better relative to other strategies than is the case in the two other models. This appears to be because, in that model, the proportion of life-years lived with immunological failure where there is also virological failure (at >1,000 copies/mL) is higher than 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 of the positive predictive values of CD4 failure for virological failure literature [29,30,31,32,33,34,35]. Thus, in the Synthesis model, the CD4 information is assumed to be a more reliable guide to viral failure than elsewhere, reducing the marginal with using viral load monitoring in that model.

Other published models besides the three used here (CEPAC and Bendavid), have also examined optimal strategies for ART monitoring, in a range of settings, and have findings that are consistent with our results[16,26]. One model, however, stands in contrast: Hamers *et al.*[27] previously suggested that viral load monitoring would be cost-saving and could improve life-expectancy. There appear to be two reasons for this discrepancy. First, the model does not model clinical or immunological failure without virologic failure; therefore the clinical and CD4 monitoring under-perform 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 adopting interventions that require additional resources means 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 GDP per capita of a country is to be "relatively cost-effective" and anything less than GDP per capita "highly cost-effective"[37]. 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 ART coverage. There may, 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 prioritization. There are also likely to be other policy goals in addition to maximizing 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[38]. 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 ART programs 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 may provide even greater benefits than our analysis of their potential lower cost implies, such as allowing more rapid results and delivery in more remote communities. Also, in settings where coverage of ART at high CD4 cell counts increases, viral load monitoring may enhance the impact that ART has on reducing HIV transmission whilst the usefulness of current CD4 monitoring algorithms may 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 may be appropriate only in wealthier countries, especially those that have scaled-up to close to full ART coverage, or if the cost of viral load testing were to fall considerably.

Author Contributions

TBH conceived the project and led the overall HIV Modelling Consortium Guidelines work. DK, PR and TBH led the co-ordination of this work and wrote the first draft of the manuscript. PR led the cost modelling and cost-effectiveness analysis. RSB, AP and NB were the principal authors contributing novel modelling and VC, JE, OK, LS assisted with contributing modelling results. DK, PR, RSB, AP, NB, AB, SW, NM, JB, VC, AC, JE, RG, AH, OK, LS, NS, AW and TBH all contributed to the design of the study, interpreting results and drafting the manuscript. MS, NM, JB and NS provided guidance on costing, economic

analysis of strategies and presentation of results. PE, MD and GH contributed to articulating the research question.

Acknowledgements

We wish to acknowledge the following persons for their participation and input over the course of the Guidelines process: Ellen McRobie, Teri Roberts, Jen Cohn, Tobias Rinke De Wit, Brooke Nichols, Theresa Rossouw, Gesine Meyer-Rath, David van de Vijver, Gert Van Zyl. VC and AP acknowledge the UCL Research Computing Services (Legion Cluster) and input to the Synthesis model from Deborah Ford, Alec Miners, Paul Revill, Fumiyo Nakagawa, and Deenan Pillay.

Role of the funding source

Funding for ALC was provided by the National Institutes of Health, including the National Institute of Allergy and Infectious Disease (NIAID) through K01 AI078754. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. PR received funding from DFID.

WHO authors contributed to design of the study, the selection of settings considered and strategies evaluated, but had no role in the development or selection of epidemiological models, conduct of the analyses or interpretation of results. The Bill and 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.

References

1. Gilks CF, Crowley S, Ekpini R, Gove S, Perriens J, Souteyrand Y et al. The WHO public health approach to antiretroviral treatment against HIV in resource-limited settings. Lancet 2006 Aug 5;368 (9534):505-10.

2. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach—2010 revision. Geneva: 2010.

3. 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(9622):1443-51. Epub 2008/04/29.

4. Estill J, Aubrière C, Egger M, Johnson LF, Wood R, Garone D, et al. 2012. Viral load monitoring of antiretroviral therapy, cohort viral load and HIV transmission in Southern Africa: a mathematical modelling analysis. AIDS 2012; 26:1403-1413.

5. Estill J, Egger M, Johnson LF, Gsponer T, Wandeler G, Davies MA et al. Monitoring of antiretroviral therapy programmes in Malawi, South Africa and Zambia: Mathematical Modelling Study. PLoS One 2013; 8(2): e57611.15.

6. Braithwaite RS, Nucifora KA, Yiannoutsous CT, Musick B, Kimaiyo S, Diero L 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.

7. Braithwaite RS, Nucifora KA, Toohey C, Kessler J, Uhler LM, Mentor SM 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).

8. Murray CJ, Acharya AK. Understanding DALYs. Journal of Health Economics 1997; 16(6), 703-730.

9. Salomon JA, Vos T, Hogan DR, Gagnon M, Naghavi M, Mokdad A 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.

10. Acharya A, Adam T, Baltussen R. Making Choices in Health: WHO Guide to Cost Effectiveness Analysis. Geneva: World Health Organization; 2004.

11. Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, USA; 2005.

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

13. Goldie SJ, Yazdanpanah Y, Losina E, Weinstein MC, Anglaret X, Walensky RP et al. Cost-effectiveness of HIV treatment in resource-poor settings—the case of Côte d'Ivoire. N Engl JMed. 2006; 355:1141–1153.

14. Bishai D, Colchero A, Durack DT. The cost effectiveness of antiretroviral treatment strategies in resource-limited settings. AIDS 2007; 21(10): 1333-1340.

15. Eaton JW, Menzies NA, Stover J, Cambiano V, Chindelevitch L, Cori A et al. How should HIV programmes respond to evidence for the benefits of earlier treatment initiation? A combined analysis of twelve mathematical models. Manuscript submitted for publication.

16. Walensky RP, Ciaranello AL, Park JE, Freedberg KA. Cost-effectiveness of laboratory monitoring in sub-Saharan Africa: a review of the current literature. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2010;51(1):85-92. Epub 2010/05/21

17. Mermin J, Ekwaru JP, Were W, Degerman R, Bunnell R, Kaharuza F, 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.

18. 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.

19. Jourdain G, Ngo-Giang-Huong N, Le Coeur S, Traisaithit S, Barbier S, Techapornroong 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; 2011; Boston, MA; 2011.

20. Saag A, Westfall A, Luhanga D, Mulenga P, Chi B, Mulenga L et al. Cluster Randomized Trial of Routine vs Discretionary Viral Load Monitoring among Adults Starting ART: Zambia. 19th Conference on Retroviruses and Opportunistic Infections; 2012; Seattle, WA; 2012.

21. Laurent C, Kouanfack C, Laborde-Balen G, Aghokeng Fobang A, Tchatchueng Mbougua JB, Boyer S, 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 Infectious Diseases 2011; 11(11): 825-833.

22. Sigaloff KCE, Hamers RL, Wallis CL, Kityo C, Siwale M, Ive P et al. Unnecessary antiretroviral treatment switches and accumulation of HIV resistance mutations; two arguments for viral load monitoring in Africa. Journal of Acquired Immune Deficiency Syndromes 2011; 58(1): 23-31.

23. Reynolds SJ, Sendagire H, Newell K, Castelnuovo B, Nankya I, Kamya M et al. Virologic versus immunologic monitoring and the rate of accumulated genotypic resistance to first-line antiretroviral drugs in Uganda. BMC Infectious Diseases 2012; 12:381.

24. 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; 23:1397–1404.

25. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N et al. Prevention of HIV-1 infection with early antiretroviral therapy. New England Journal of Medicine 2011; 365: 493-505.

26. Kimmel AD, Weinstein MC, Anglaret X, Goldie SJ, Losina E, Yazdanpanah Y, 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(3):258-68. Epub 2010/04/21.

27. 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(13):1663-72. Epub 2012/06/15.

28. Keiser O, Tweya H, Braitstein P, Dabis F, MacPhail P, Boulle A, et al. Mortality after failure of antiretroviral therapy in sub-Saharan Africa. Tropical Medicine and International Health 2010; 15(2): 251-258.

29. 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 Inf. Dis. 2008; 8:89.

30. 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. Inf. Dis. 2007; 11: 413-416.

31. Keiser O, McPhail P, Boulle A et al. Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy. Trop Med Intl Health 2009; 14(10): 1220-1225.

32. Kiragga AN, Castelnuovo B, Kamya MR, Moore R, Manabe YC. Regional differences in predictive accuracy of WHO immunologic failure criteria. AIDS 2012; 26(6): 768-770.

33. 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(15): 1971-1977.

34. 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(5): 477-484.

35.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.

36. Boyer S, March L, Kouanfack C, Laborde-Balen G, Marino P, Aghokeng Fobang A, 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 Infectious Diseases; available online 18 April 2013: http://dx.doi.org/10.1016/S1473-3099(13)70073-2.

37. Choosing interventions that are cost-effective: Cost-effectiveness thresholds [homepage on the Internet]. Geneva: World Health Organization; 2012 [cited 30 April 2012]. Available from: http://www.who.int/choice/costs/CER_thresholds/en/index.html

38. Walensky RP, Wood R, Ciaranello AL, Paltiel AD, Lorenzana SB, et al. Scaling Up the 2010 World Health Organization HIV Treatment Guidelines in Resource-Limited Settings: A Model-Based Analysis. PLoS Med 2010; 7(12): e1000382. doi:10.1371/journal.pmed.1000382