Clinical Evaluation of the Potential Utility of Computational Modeling as an HIV Treatment Selection Tool by Physicians with Considerable HIV Experience

Brendan A. Larder, Ph.D.,[1] Andrew Revell, Ph.D.,[1] JoAnn M. Mican, M.D.,[2] Brian K. Agan, M.D.,[3]
Marianne Harris, M.D.,[4] Carlo Torti, M.D.,[5] Ilaria Izzo, M.D.,[5] Julia A. Metcalf, B.A.,[2]
Migdalia Rivera-Goba, R.N.,[6] Vincent C. Marconi, M.D.,[3] Dechao Wang, Ph.D.,[1] Daniel Coe, B.Sc.,[1]
Brian Gazzard, M.D.,[7] Julio Montaner, M.D.,[4] and H. Clifford Lane, M.D.[2]

Abstract

The HIV Resistance Response Database Initiative (RDI), which comprises a small research team in the United Kingdom and collaborating clinical centers in more than 15 countries, has used antiretroviral treatment and response data from thousands of patients around the world to develop computational models that are highly predictive of virologic response. The potential utility of such models as a tool for assisting treatment selection was assessed in two clinical pilot studies: a prospective study in Canada and Italy, which was terminated early because of the availability of new drugs not covered by the system, and a retrospective study in the United States. For these studies, a Web-based user interface was constructed to provide access to the models. Participating physicians entered baseline data for cases of treatment failure and then registered their treatment intention. They then received a report listing the five alternative regimens that the models predicted would be most effective plus their own selection, ranked in order of predicted virologic response. The physicians then entered their final treatment decision. Twenty-three physicians entered 114 cases (75 unique cases with 39 entered twice by different physicians). Overall, 33% of treatment decisions were changed following review of the report. The final treatment decisions and the best of the RDI alternatives were predicted to produce greater virologic responses and involve fewer drugs than the original selections. Most physicians found the system easy to use and understand. All but one indicated they would use the system if it were available, particularly for highly treatment-experienced cases with challenging resistance profiles. Despite limitations, the first clinical evaluation of this approach by physicians with substantial HIV-experience suggests that it has the potential to deliver clinical and economic benefits.

Introduction

THE CURRENT OBJECTIVE of combination antiretroviral therapy for HIV infection is long-term suppression of circulating virus to below the limit of detection of the assays currently in widespread use, typically 40 or 50 copies of HIV RNA per milliliter.[1,2] Despite the availability in well-resourced health care settings of approximately 25 antiretroviral drugs, treatments continue to fail, often with the emergence of drug-resistant virus, necessitating a change in therapy. Resuppressing the virus and maintaining it at low levels for the lifetime of the patient requires careful choices to overcome drug resistance, stay one step ahead of viral evolution, minimize toxicity, and facilitate the patient's adherence. Faced by the sheer complexity of resistance, the number of potential drug combinations available and the competing clinical and economic considerations affecting the treatment decision, the intelligent, individualised sequencing of antiretroviral therapy is highly challenging.[3] For physicians with limited experience or resources, antiretroviral treatment decision-making can become even more difficult.

The standard of care in well-resourced settings is to monitor the patient's viral load regularly, with detection of an increase often triggering a change of antiretroviral drug

---

[1]The HIV Resistance Response Database Initiative (RDI), London, United Kingdom.
[2]National Institutes of Allergy and Infectious Diseases, Bethesda, Maryland.
[3]Infectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, Maryland.
[4]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada.
[5]University of Brescia, Brescia, Italy.
[6]National Institutes of Health, Clinical Center, Bethesda, Maryland.
[7]Chelsea and Westminster Hospital, London, United Kingdom.

therapy. A genotypic resistance test is usually ordered to identify the point mutations that may confer drug resistance. The interpretation of the genotype is complex and is usually performed using rules-based interpretation software that relates point mutations to single drug susceptibility.[4] However, there is no gold standard interpretation system: different systems provide different interpretations and do not always concur well.[5–9] Moreover, it is difficult to relate the predicted susceptibility to individual drugs to the likely relative responses to potential drug combinations: raw genotype sensitivity scores being significant but relatively weak predictors of virologic response.[11–14] Bioinformatics have been used most commonly to predict phenotype from genotype and then relate a cutoff in predicted phenotype to a categorical response.[15,16] Again, it is difficult to relate this categorical prediction for an individual drug to the relative responses that may be achieved with different candidate combinations.

Models that provide a quantitative prediction of virologic response to combination therapy, rather than to individual drugs, directly from the genotype and other information may offer a potential clinical advantage. However, this can be challenging given that a very large dataset is required to accommodate multiple possible drug-genotype permutations with their respective drug response data.[17] The HIV Resistance Response Database Initiative (RDI) was established in 2002 to be the global repository for data, collected from clinical practice around the world, required to develop such models.[18]

Currently we have collected data from approximately 65,000 patients. We have previously trained computational models using subsets of these data to predict virologic treatment response from genotype, viral load, CD4 cell count, and treatment history. When tested with independent retrospective data, the models have proved accurate, with correlations between the predicted and actual changes in viral load (ΔVL) in excess of 0.8 ($r^2 \geq 0.65$).[19] More recently, models trained to predict the probability of a regimen reducing the viral load to less than 50 copies per milliliter have achieved accuracy and an area under the receiver operating characteristic (ROC) curve (AUC) of 0.80 and above, which compares favorably with the AUC values of 0.60–0.65 typically achieved by common rules-based genotype interpretation systems.[11,20] In addition the models are able to identify alternative combinations of antiretroviral drugs that are predicted to be effective for a substantial proportion of cases that resulted in virologic failure in the clinic and that were predicted to fail by the models.[21]

These results suggest that computational models may have a useful role as an aid to antiretroviral treatment decision-making. Before making such models available, however, it is clearly important to test their potential utility in clinical practice and collect input from HIV physicians with regard to the design of the interface through which such models could be accessed.

Here we report results from two complementary open clinical pilot studies of the use of computational models embedded in a prototype Web-based treatment decision-making system. Study 1 was prospective and Study 2 used retrospective data. The common objectives of the studies, which were designed to run in parallel addressing the same research questions, were to evaluate the proportion of HIV physicians that might modify their treatment decisions as a result of using the system (to inform the powering of subsequent controlled trials), to obtain a qualitative assessment of the ease of use of the Web-based interface, and to obtain suggestions from experienced HIV physicians for its improvement.

Methods

The basic package of information that is used for the training of the RDI's models is the treatment change episode (TCE). This comprises key information required by the models from a patient who has had a new antiretroviral treatment started, in order to develop a prediction of virologic response. It includes baseline genotype, viral load, CD4$^+$ T-lymphocyte (CD4) count, treatment history and time to follow-up as well as the follow-up viral load value: the response variable that the models are trained to predict. For this study, a random forest model and a committee of 10 artificial neural network models were trained to predict the change from baseline viral load (ΔVL) using data from 3188 TCEs from multiple clinical sources, using methodology previously described.[15,16] The 82 input variables used to train the models were 58 mutations in the HIV RNA regions encoding protease and reverse transcriptase, 17 antiretroviral drugs (zidovudine, didanosine, stavudine, abacavir, lamivudine [3TC]/emtricitabine [FTC], tenofovir disoproxil fumarate [DF], efavirenz, nevirapine, indinavir, nelfinavir, ritonavir, saquinavir, (fos)amprenavir, lopinavir, atazanavir, darunavir, and enfuvirtide [T20]), viral load, CD4 count, treatment history (4 variables) and time to follow-up. The predictions of the models were combined (arithmetic mean) and the performance of the system tested using an independent set of 100 TCEs. The models' predictions correlated with the actual ΔVL values of the test TCEs with a coefficient of 0.83 ($r^2 = 0.68$) and a mean absolute difference of 0.49 log$_{10}$ copies per milliliter.[16]

An online treatment decision tool was developed through which physicians in the participating centers could access the models described above via an interface on the RDI website using password-protected user accounts. Data, including all the input variables required by the models as described above from patients requiring a treatment change (viral load > 3 log$_{10}$ copies per milliliter on current antiretroviral regimen) were entered by the participating physicians online, together with the physician's intended next treatment and any contraindicated drugs that the physician wanted excluded from the modeling. These baseline data were delivered to the computational models, which produced predictions of 12-week virologic response to the regimen that the physician intended to use plus alternative regimens that are in clinical use (as determined from the RDI database), excluding any drugs that the physician had ruled out for toxicity or other reasons. The system produced a pdf report listing the baseline data plus the predictions of responses for the physician's treatment intention and the five regimens predicted by the models to be most effective from the range of alternatives in the following three categories:

1. No more drugs* than the physician's selection
2. No more than six drugs*
3. No more than six drugs* including any contraindicated drugs *excluding ritonavir as a booster of protease inhibitors

Regimens were listed on the report in order of predicted ΔVL, grouped into 0.5 $\log_{10}$ bands. When regimens with different numbers of drugs were predicted to produce the same virologic response, precedence was given to the simpler regimen. The physician's original selection was listed among the RDI alternatives in the rank position corresponding to the models' prediction of ΔVL. Having reviewed the report, the physicians were then required to enter their final treatment decision. The 12-week virologic response to this regimen was modeled and the prediction stored.

Following completion of their cases, participating physicians filled-out an on-line evaluation questionnaire in which they were asked to rate the following attributes of the system:

- The ease of use of the Web-based interface.
- The ease with which the RDI report could be understood.
- The utility of the system in helping treatment decision-making.
- The anticipated frequency with which they would use the system if it were freely available.

Finally, they were invited to make suggestions for improvement of the system and indicate for what type of patients they would be most likely to use the system.

The following outcome data were collected and analyzed in both studies:

1. The proportion of treatment decisions that were changed following review of the predictions on the RDI report.
2. Evaluation of the system:
   a. Ratings of the ease of use, utility and predicted frequency of use of the system
   b. Evaluation of the number of drugs and the nature of the regimens selected by the physicians before and then after review of the RDI predictions

Participating centers and physicians

The physicians involved in these studies had considerable experience in the management of patients with HIV and AIDS. Study 1, the prospective pilot study was run in two centers, in Canada (BC Centre for Excellence in HIV/AIDS) and Italy (University of Brescia). The participating clinicians were the principal investigators: Dr. Julio Montaner, with more than 20 years' experience and Dr. Carlo Torti, with more than 14 years experience treating patients with HIV and AIDS. Both run clinics with many hundreds of patients. Study 2, the retrospective study, was conducted in the National Institutes of Allergy and Infectious Diseases (NIAID) HIV clinic on the Bethesda, Maryland, campus. The physicians who took part were volunteers from this clinic, the majority of whom had more than 10 years' experience treating patients with HIV and AIDS. All had treated more than 20 patients in the previous 2 years.

Patient/case recruitment for the two studies

In Study 1, patients requiring a treatment change were recruited in real time and underwent informed consent. Twelve weeks after the new regimen was started, the physician obtained and entered the follow-up plasma viral load value and completed an online evaluation questionnaire. The target for recruitment was 150 patients.

In Study 2, clinical data from patients in the participating HIV clinics whose treatment had been changed in the past were stripped of identifiers and information relating to the new regimen and allocated at random to physicians who had not been involved in the treatment of the patients. These data were treated as if they were from new cases of treatment failure and entered in to the system as described above. The target was for 100 cases to be used.

Results

Patient disposition

Study 1: The prospective pilot. Ten male patients had completed the study when it was terminated because of the availability of new drugs (raltegravir, etravirine, and maraviroc) not included in the system. The patients had a mean baseline plasma viral load on failing therapy of 4.1 $\log_{10}$ copies per milliliter and a mean CD4 count of 367 cells/mm^3. They had previous exposure to a mean of 4.4 nucleoside or nucleotide reverse transcriptase inhibitors (NRTIs) and 2.2 protease inhibitors (PIs) with six having previous exposure to a single non-nucleoside reverse transcriptase inhibitor (NNRTI). The mean numbers of mutations relating to NRTIs, NNRTIs, and PIs were 4.7, 0.5, and 4.7, respectively.

Study 2: The retrospective pilot. Sixty-five patients who had had an antiretroviral treatment change in the past were identified from the NIAID and DoD clinics. These cases were allocated to 21 physicians. In order to achieve the target sample size, 39 of the cases, selected at random, were processed twice by two different physicians, to give a total of 104 cases entered. The patients had a mean baseline plasma viral load on failing therapy of 4.75 $\log_{10}$ copies per milliliter and a mean CD4 count of 322 cells/mm^3. They had previous exposure to a mean of 3.9 NRTIs, 1.02 NNRTIs, and 2.6 PIs. The mean numbers of mutations relating to NRTIs, NNRTIs and PIs were 3.3, 1.2 and 4.5, respectively.

Outcomes

In Study 1, the physicians changed their decision following the receipt and review of the RDI report in 5 of 10 cases (50%). In Study 2 the treatment decision was changed in 33 of 104 cases (32%). Overall, treatment decisions were changed in 38 of 114 cases or 33%. Based on this, the estimate of the proportion of treatment decisions that would be changed if the RDI system were in use is 33%, $\pm$ 8.6% (24.4–41.6%) at the 95% confidence level. The treatment decision was changed to a regimen listed on the RDI report in 5 of these 38 cases but was amended further in the remaining 33 cases for a variety of reasons, including tolerability, physician judgment, patient preference, and an intention to use new drugs.

The first of the secondary outcomes was the physicians' evaluation of the system. The physicians participating in the studies (two in Study 1 and 21 in Study 2) completed an evaluation questionnaire covering the ease of use of the system, its utility as a tool to help with treatment decision-making and the predicted frequency of use if it were freely available on the Internet. Each question was followed by five response options. The questions, response options, and results are presented in Table 1.

Twenty (87%) participating physicians found it easy or very easy to enter their baseline data, two (9%) found it satisfactory and one (4%) registered it as quite difficult. Two thirds (15 or 65%) of participants found the Web interface easy or very easy to use overall. Five more found it satisfactory and three found it difficult or very difficult. All 23 participating physicians found the report at least satisfactory in terms of ease of comprehension, with approximately three quarters (74%) finding it easy or very easy to understand. Five (22%) of the participants found the system quite useful and 11 (48%) satisfactory in terms of its utility in making a treatment decision. Seven (30%) participants did not find the system very useful. One participant (4%) indicated that they would use the system very frequently, 7 (30%) quite frequently, 11 (48%) sometimes, 3 (13%) infrequently, and 1 (4%) never.

The physicians were also asked to suggest improvements for the system and indicate the types of cases for which they would use the system. Responses made by more than one physician are listed in Table 2.

A comparison of the models' predictions for physicians' initial and final selections and the alternative regimens predicted to be most effective

In this analysis, the system's predictions for the physicians' original treatment selections, the best of the RDI alternatives and the physicians' final selections were compared with one another. The predicted mean and median changes in viral load from baseline are listed in Table 3.

The results show a clear pattern with the worst virologic responses predicted to result from the physicians' original treatment choices, somewhat better responses predicted for their final decisions, and the best responses predicted for the top RDI alternatives. If only those cases where the treatment decision was changed are considered, the difference between the predicted responses to the original and final selections becomes more marked. For example, the proportions of cases predicted to achieve a reduction in viral load of at least $2 \log_{10}$ copies per milliliter or more were 39% for the physicians' original choices, 50% for their final selections, and 58% for the best of the RDI alternatives.

The predicted virologic responses for the best alternatives on the RDI report were significantly greater than those for the physicians' initial treatment selections. The predicted responses for the physicians' final treatment selections showed a nonsignificant trend toward being greater than for their

initial decisions. When the subset of 38 cases where the treatment decision was changed was examined, both the physicians' final treatment decisions and the best alternatives were predicted to produce significantly greater virologic responses than the physicians' initial selections.

Analysis of antiretroviral treatments used, selected by the physicians and predicted by the RDI system to be most effective

The numbers of drugs in the physicians' initial and final treatment decisions and in the regimen predicted to give the largest virologic response on the RDI report (Section A) are presented in Table 4. The mean number of drugs for each category was 3.72, 3.66, and 3.36, respectively. The top RDI alternatives involved a mean of 0.36 fewer drugs per regimen than the physicians' original selections. The proportion of regimens containing no more than three drugs increased from 41% in the physicians' original selections to 46% in their final selections and 70% for the top RDI alternatives from Section A of the report.

When only those cases in which a change in treatment decision occurred were analyzed, the reduction in the mean number of drugs in the final regimen was −0.13. The difference between the original treatment decision and the best of the RDI alternatives was −0.34 drugs. The proportion of regimens containing no more than 3 drugs increased from 32% in the physicians' original selections to 42% in their final selections and 58% for the top RDI alternatives. The mean numbers of drugs for cases where treatments were changed in Study 1, the prospective study, were 4.4, 3.8 and 3.6, a reduction of 0.6 and 0.8 drugs per regimen for the physician's final decision and the top RDI recommendation, respectively.

An analysis of the different categories (by drug class) of regimen involved in the physicians' initial intentions and final decisions and the top RDI predictions are summarised in Table 5 and revealed the following patterns:

- More regimens involving a PI and two NRTIs in the RDI top alternatives than were selected by physicians.
- No examples of a regimen consisting of a single NNRTI and two NRTIs as the top regimen on an RDI report compared to 11% of physicians' selections
- A higher frequency of triple class therapy (NRTI, NNRTI and a PI) on the RDI reports (14%) compared with only 2%–4% of physicians' selections.

Table 1. Summary of Physician's Evaluations of the System (Most Popular Responses Highlighted)

Question	Response options with number (%) of the 23 physicians selecting each option				
How easy was it to enter the baseline data?	Very easy 6 (26%)	**Quite easy 14 (61%)**	Satisfactory 2 (9%)	Quite difficult 1 (4%)	Very difficult 0 (0%)
How easy to use was the Web-based interface overall?	Very easy 6 (26%)	**Quite easy 9 (39%)**	Satisfactory 5 (22%)	Quite difficult 1 (4%)	Very difficult 2 (9%)
How easy was it to understand the RDI report?	Very easy 6 (26%)	**Quite easy 11 (48%)**	Satisfactory 6 (26%)	Quite difficult 0 (0%)	Very difficult 0 (0%)
How useful was the system in making treatment decisions?	Very useful 0 (0%)	Quite 5 (22%)	**Satisfactory 11 (48%)**	Not very 7 (30%)	Not at all 0 (0%)
If the system were freely available, how frequently would you use it?	Very 1 (4%)	Quite 7 (30%)	**Sometimes 11 (48%)**	Infrequent 3 (13%)	Never 1 (4%)

TABLE 2. MOST COMMON UNPROMPTED RESPONSES (MADE BY MORE THAN ONE PHYSICIAN)

How could baseline data entry have been made easier?	Allow direct date entry rather than drop-down menus (4)
	Enable user to save incomplete baseline data for completion at a later time/date (2)
Are there any other types of information that you would have liked on the report?	Allow concerns about use of certain meds even though there are not absolute contraindications (2)
How could the system have been made more useful as a treatment decision aid?	Include newer drugs (5)
	Exclude older drugs with toxicity issues, e.g. ddI and d4T (3)
For what types of patient would you be most likely use the system?	Highly treatment-experienced patients (6)
	Patients with extensive resistance patterns (6)
	Complex resistance patterns e.g., PI and NNRTI resistance (3)

PI, protease inhibitor; NNRTI, non-nucleotide reverse transcriptase inhibitor.

- A higher frequency of triple class therapy including enfuvirtide among the physicians' original selections (32%) than in their final decisions (24%) and the RDI reports (23%).

Discussion

The results of these clinical studies indicate that the proportion of treatment decisions made by experienced HIV physicians (using comprehensive clinical and laboratory monitoring including viral loads, CD4 counts and genotypes) for treatment-experienced patients on failing therapy that would be changed if the RDI system were in use is 33% ($\pm 8.6\%$ at the 95% confidence level). This may be a conservative estimate for the following reasons. First, the physicians in this study registered their initial treatment decision before receiving the RDI report. They subsequently had to register their changed decision online, thereby "admitting" that their original decision might be improved upon. Second, the system did not cover several of the newer drugs becoming available at the time (raltegravir, etravirine, maraviroc). Finally, the models used in this study and therefore the suggestions of alternatives on the reports provided to the physicians, included older drugs that were being phased out of clinical practice for tolerability issues, including stavudine.

It is noteworthy that in the majority of cases where the treatment decision was changed, the new regimen selected was not one taken directly from the RDI report but one or more further drug changes were made, for reasons of tolerability, physician judgment, patient preference, and a desire to use new drugs. While all of the alternative regimens modeled by the system were in clinical use, some (for example triple class therapy) were considered unusual and rejected for this reason. Restriction of the lists of alternative regimens to be modeled to those that are in most common use could be considered but this could rule out unusual but potentially effective alternatives for those patients with limited treatment options. In addition, the regimens that were finally selected by the physicians were predicted by the models to do relatively well, but they did not make the top five presented on the report. The expansion of the report to include more than five regimens could be considered. Most importantly, the inclusion of the newest drugs in the models would clearly improve the utility of the system. Notwithstanding the fact that most final treatment decisions were not exactly one of the top five alternatives on the report, review of the report led the physicians to reflect on their treatment decisions and the final decisions reached were improved as a result—with fewer drugs and superior predicted responses.

The RDI system was demonstrably easy to use and the report easily understood by most users. The great majority of participants found it easy or very easy to enter their baseline data, the most onerous part of the user's interaction with the system, with 96% rating it as satisfactory or better. The one

TABLE 3. COMPARISON OF PREDICTED CHANGE IN VIRAL LOAD FROM BASELINE

	Physician's selection		RDI system	
	Initial	Final	A[a]	B[a]
A: All cases (n = 114)				
Mean predicted change in viral load from baseline	−1.88	−1.90	−2.03	−2.03
Median	−1.85	−1.85	−1.98	−1.98
Proportion with >2 log reduction	39%	41%	50%	50%
Statistical significance (vs physician's initial selection)[b]		$p = 0.06$	$p < 0.0001$	$p < 0.0001$
B: Cases where the treatment decision was changed (n = 38)				
Mean	−1.92	−1.99	−2.12	−2.13
Median	−1.91	−1.99	−2.06	−2.07
Proportion with >2 log reduction	39%	50%	58%	58%
Statistical significance (vs. physician's initial selection)[b]		$p < 0.05$	$p < 0.0001$	$p < 0.0001$

[a]A, Alternative regimens with no more drugs than the physician's initial selection; B, Alternative regimens with no more than six drugs.
[b]One tailed t tests.

TABLE 4. NUMBER OF DRUGS IN EACH REGIMEN AND DIFFERENCES (Δ) FROM THE PHYSICIANS' ORIGINAL INTENTION

	Intended		Final decision			RDI top		
	n	%	n	%	Δ	n	%	Δ
All cases								
3 drugs	47	41%	53	46%	+5%	80	70%	+29%
4 drugs	52	46%	48	42%	−4%	27	24%	−22%
5 drugs	15	13%	12	11%	−2%	7	6%	−7%
6 drugs	0	0%	1	1%	+1%	0	0%	0
Total cases	114		114			114		
Total drugs	424		417		−7	383		−41
Mean drugs/regimen	3.72		3.66		−0.06	3.36		−0.36
Statistical significance (vs. the physician's original intention					ns			$p < 0.05$
Cases where treatment decision changed								
3 drugs	12	32%	16	42%	+10%	22	58%	+28%
4 drugs	19	50%	17	45%	−5%	12	32%	−18%
5 drugs	7	18%	4	10%	−8%	4	11%	−7%
6 drugs	0	0%	1	3%	+3%	0	0%	0%
Total cases	38		38			38		
Total drugs	147		142		−5	134		−13
Mean drugs/regimen	3.87		3.74		−0.13	3.53		−0.34
Statistical significance (vs. the physician's original selection					ns			$p < 0.05$

ns, not significant.

physician who found it difficult encountered initial problems accessing the system from within the virtual private network in their clinic, because of security issues, but did not report any difficulties in using the system once this was resolved and the system accessed. The most common suggestion for improvement to the user interface was to allow direct entry of the dates for baseline data rather than using drop-down menus. Two participants suggested enabling the user to save incomplete baseline data. This slows data entry down somewhat but would enable the user to return to complete inputting the data. These suggestions are being implemented.

There were no reports of any difficulty understanding the RDI report and almost three quarters of participating physicians found it easy or very easy to understand. Two participants suggested adding warnings over drugs that do not constitute absolute contraindications. In response to this, the new system under development will enable users to rule out

TABLE 5. DIFFERENT CATEGORIES OF REGIMEN

	Initial selection		Final selection		RDI top	
PI + 2 NRTI	36	32%	42	37%	59	52%
PI + 3 NRTI	17	15%	25	22%	10	9%
NNRTI + 2 NRTI	13	11%	12	11%	0	0%
NNRTI + 3 NRTI	4	4%	5	5%	0	0%
3 class (NRTI, NNRTI, PI)	5	4%	2	2%	16	14%
3 class incl enfuvirtide	37	32%	27	24%	26	23%
≥ 4 class	1	1%	1	1%	0	0%
Other	1	1%	0	0%	3	3%

PI, protease inhibitor; NRTI, nucleoside or nucleotide reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitor.

drugs from the modeling for any reason including issues of access, concerns over tolerability or, for example, physician or patient preferences.

The majority of participants found the system satisfactory or better in terms of its usefulness in making a treatment decision. Nevertheless, seven participants did not. Four of these had quite straightforward cases in terms of resistance—two of these physicians explicitly commented that this was why they did not find the system useful. The remaining five physicians all cited the need to update the drugs/regimens used in the system as the reason for its limited utility.

All but one of the physician participants indicated that they would use the system, one third of them frequently or very frequently and approximately half sometimes. The one participant who indicated that they would not use the system had cases with very straightforward genotypes and their main suggestion for improvement was the inclusion of newer drugs. By far the most frequent descriptions of the type of patient that participants would use the system for were highly treatment-experienced patients (six physicians) and those with extensive or complex resistance patterns (nine physicians).

Having established that the system is easy to use, likely to change treatment decisions and to be used in practice it is important to examine whether such changes might be beneficial to patient care. We have previously established the accuracy of the models in predicting virologic response and the results of this study demonstrate that the predicted responses for the physicians' final treatment selections were significantly better than for their original selections in cases where the treatment decision was changed. More marked was the difference between the prediction for the physicians' original treatment selections and the best of the alternatives presented on the RDI report. This was highly

significant across all the cases. While some of the differences in the predicted virologic responses observed were modest, the models appear to be at least as good as, if not better than HIV-experienced physicians in identifying effective regimens, suggesting that the system could be of benefit not only for difficult cases with complex patterns of resistance, but more generally for less experienced physicians. The number of drugs in the final regimen was reduced slightly (−0.13) and not significantly from the number in the physician's initial selection. The number of drugs in the best of the RDI alternatives, however, was more substantially and statistically significantly reduced, by approximately one third of a drug. In Study 1, the prospective study, the reduction in the number of drugs was particularly marked at −0.6 and −0.8 drugs per regimen for the physician's final decision and the top RDI recommendation respectively. These results, considered together with the fact the physicians switched to one of the RDI alternatives in only 13% of cases where the decision was changed, suggest that more direct use of the system in a prospective setting has the potential to produce greater benefits in terms of virologic responses and the reduction in the number of drugs used than was observed here.

The studies reported here have a number of limitations. The first is that Study 2 used retrospective cases. The physicians allocated these cases were, therefore, not making real treatment decisions for their patients in real time, which may have affected their decision-making. It also prevented any follow-up after the final treatment decisions had been made.

There were two significant limitations to the models in this system. First, as already discussed they were developed at a time when there were insufficient long-term follow-up data on etravirine, raltegravir, and maraviroc (all of which were being used with increasing frequency at the time the studies were recruiting) to include these drugs in the models' predictions. This resulted in the early termination of the prospective pilot study. Another limitation of the models is that they were trained to predict the change in viral load from baseline, which reflected the fact that most heavily pretreated patients at the time that the data were collected for training the models were not achieving complete virologic suppression. However, when these studies were run, this situation had changed and the great majority of patients were achieving viral suppression following a treatment change. This may have limited the apparent differences between the predicted responses for the physicians' initial and final choices and the RDI alternatives. The RDI has since developed models that predict the probability of achieving a viral load below a certain limit, for example 50 copies HIV RNA per milliliter, which is in line with the current objective of HIV treatment.

Notwithstanding these limitations, this is the first clinical evaluation of this approach and the encouraging results indicate that the system is easy to use and has the potential to provide significant benefits in terms of the simplicity and acceptability of therapy, the virologic response to that therapy and its cost. The results indicate that further development and clinical study of the RDI system is warranted. In the meantime, an updated version of the system incorporating suggestions made by physicians in these studies is being made available as an experimental tool and a version that does not require genotypic resistance information is in development for resource-limited settings where these assays are not available.[21]

Author Disclosure Statement

No competing financial interests exist.

References

1. Thompson MA, Aberg JA, Cahn C, et al. Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society—USA Panel. JAMA 2010;304:321–333.
2. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1–infected adults and adolescents. Department of Health and Human Services. December 1, 2009;1–161. www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf (Last accessed July 30, 2010).
3. Hirsch MS, Günthard HF, Schapiro JM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society—USA panel. Clin Infect Dis 2008;47:266–285.
4. Sturmer M, Doerr HW, Preiser W. Variety of interpretation systems for human immunodeficiency virus type 1 genotyping: Confirmatory information or additional confusion? Curr Drug Targets Infect Disord 2003;3:373–382.
5. Schapiro JM, De Luca A, Harrigan R, et al. Resistance assay interpretation systems vary widely in method and approach. Antivir Ther 2001;6(Suppl 1):131.
6. Shafer RW, Gonzales MJ, Brun-Vezinet F. Online comparison of HIV-1 drug resistance algorithms identifies rates and causes of discordant interpretations. Antivir Ther 2001;6:101.
7. Torti C, Quiros-Roldan E, Keulen W, et al. For the GenPherex Group of the MaSTeR Cohort. Comparison between rules-based human immunodeficiency virus type 1 genotype interpretations and real or virtual phenotype: Concordance analysis and correlation with clinical outcome in heavily treated patients. J Infect Dis 2003;188:194–201
8. Sturmer M, Doerr HW, Staszewski S, Preiser W. Comparison of nine resistance interpretation systems for HIV-1 genotyping. Antivir Ther 2003;8:239–244.
9. De Luca A, Cingolani A, Di Giambenedetto S, et al. Variable prediction of antiretroviral treatment outcome by different systems for interpreting genotypic human immunodeficiency virus type 1 drug resistance. J Infect Dis 2003;15:1934–1943
10. DeGruttola V, Dix L, D'Aquila R, et al. The relation between baseline HIV drug resistance and response to antiretroviral therapy: Re-analysis of retrospective and prospective studies using a standardized data analysis plan. Antivir Ther 2000;5:41–48.
11. Frentz D, Boucher CAB, Assel M, et al. Comparison of HIV-1 genotypic resistance test interpretation systems in predicting virological outcomes over time. PLoS One 2010;5:e11505.
12. Gallego O, Martin-Carbonero L, Aguero J, de Mendoza C, Corral A, Soriano V. Correlation between rules-based interpretation and virtual phenotype interpretation of HIV-1 genotypes for predicting drug resistance in HIV-infected individuals. J Virol Methods 2004;121:115–118.
13. Larder BA, Revell A, Wang D, Harrigan R, Montaner J, Wegner S. Neural networks are more accurate predictors of virological response to antiretroviral therapy than rules-based genotype interpretation systems [Abstract 653]. 13th Conference on Retroviruses and Opportunistic Infections. February 5–8, 2006.
14. Larder BA, Wang D, Revell A. Application of artificial neural networks for decision support in medicine. In: Livingstone DS, ed. Artificial Neural Networks: Methods and Applications. New York: Humana Press (Springer-Verlag), New York, 2008, 119–132.
15. Beerenwinkel N, Daumer M, Oette M, et al. Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res 2003;31:3850–3855.
16. Beerenwinkel N, Sing T, Lengauer T, et al. Computational models for the design of effective therapies against drug resistant HIV strains. Bioinformatics 2005;21:3943–3950.
17. DiRienzo G, DeGruttola V. Collaborative HIV resistance-response database: Sample size for detection of relationships between HIV-1 genotype and HIV-1 RNA response using a non-parametric approach. Antivir Ther 2002;7:S71.
18. Larder BA, DeGruttola V, Hammer S, et al. The international HIV resistance response database initiative: A new global collaborative approach to relating viral genotype treatment to clinical outcome. Antivir Ther 2002;7:S84.
19. Larder BA, Wang D, Revell A, et al. The development of artificial neural networks to predict virological response to combination HIV therapy. Antivir Ther 2007;12:15–24.
20. Larder BA, Wang D, Revell AD, et al. The development of computational models that accurately predict virological response to HIV therapy to power an online treatment selection tool. Antivir Ther 2010;15:A89.
21. Revell AD, Wang D, Harrigan R, et al. Modelling response to HIV therapy without a genotype—An argument for viral load monitoring in resource-limited settings. J Antimicrob Chemother 2010;65:605–607.

Address correspondence to:
Andrew Revell, Ph.D.
RDI
14 Union Square London N1 7DH
United Kingdom

E-mail: andrewrevell@hivrdi.org