Journal of Antimicrobial Chemotherapy (2009) **64**, 616–624
doi:10.1093/jac/dkp252
Advance Access publication 19 July 2009

Rules-based HIV-1 genotypic resistance interpretation systems predict 8 week and 24 week virological antiretroviral treatment outcome and benefit from drug potency weighting

Maurizio Zazzi[1]*, Mattia Prosperi[2,3], Ilaria Vicenti[1], Simona Di Giambenedetto[4], Annapaola Callegaro[5], Bianca Bruzzone[6], Fausto Baldanti[7], Angela Gonnelli[8], Enzo Boeri[9], Elisabetta Paolini[10], Stefano Rusconi[11], Andrea Giacometti[12], Franco Maggiolo[13], Stefano Menzo[14] and Andrea De Luca[4] on behalf of the ARCA Collaborative Group†

[1]Department of Molecular Biology, University of Siena, Siena, Italy; [2]INMI Lazzaro Spallanzani, Rome, Italy; [3]Informa SRL, Rome, Italy; [4]Institute of Clinical Infectious Diseases, Catholic University 'Sacro Cuore', Rome, Italy; [5]Microbiology and Virology Laboratory, Ospedali Riuniti, Bergamo, Italy; [6]Microbiology and Virology Laboratory, San Martino Hospital, Genoa, Italy; [7]Unit of Virology, IRCCS San Matteo Hospital, Pavia, Italy; [8]Division of Infectious Disease, Siena University Hospital, Siena, Italy; [9]Diagnostica & Ricerca San Raffaele, Milan, Italy; [10]Unit of Immunohematology and Transfusional Medicine, Cremona Hospital, Cremona, Italy; [11]Section of Infectious Diseases and Immunopathology, Department of Clinical Sciences 'Luigi Sacco', Milan, Italy; [12]Institute of Clinical Infectious Diseases, Ancona University Hospital, Ancona, Italy; [13]Division of Infectious Diseases, Ospedali Riuniti, Bergamo, Italy; [14]Institute of Microbiology and Biomedical Sciences, Marche Polytechnical University, Ancona, Italy

Received 16 April 2009; returned 1 June 2009; revised 15 June 2009; accepted 21 June 2009

***Objectives*: To test retrospectively the ability of four freely available rules-based expert systems to predict short- and medium-term virological outcome following an antiretroviral treatment switch in pre-treated HIV-1 patients.**

***Methods*: The HIV-1 genotype interpretation systems (GISs) HIVdb, ANRS, Rega and AntiRetroScan were tested for their accuracy in predicting response to highly active antiretroviral therapy using 8 week ($n=765$) and 24 week ($n=634$) follow-up standardized treatment change episodes extracted from the Italian Antiretroviral Resistance Cohort Analysis (ARCA) database. A genotypic sensitivity score (GSS) was derived for each genotype–treatment pair for the different GISs and tested as a predictor of virological treatment outcome by univariable and multivariable logistic regression as well as by receiver operating characteristic curve analysis. The two systems implementing drug potency weights (AntiRetroScan and Rega) were evaluated with and without this correction factor.**

***Results*: All four GSSs were strong predictors of virological treatment outcome at both 8 and 24 weeks after adjusting for baseline viro-immunological parameters and previous drug exposure (odds ratios ranging from 2.04 to 2.43 per 1 unit GSS increase; $P<0.001$ for all the systems). The accuracy of AntiRetroScan and Rega was significantly increased by drug potency weighting with respect to the unweighted versions ($P\leq0.001$). HIVdb and ANRS also increased their performance with the same drug potency weighting adopted by AntiRetroScan and Rega, respectively ($P<0.001$ for both analyses).**

***Conclusions*: Currently available GISs are valuable tools for assisting antiretroviral treatment choices. Drug potency weighting can increase the accuracy of all systems.**

Keywords: genotype, drug resistance, algorithm

*Corresponding author. Section of Microbiology, Department of Molecular Biology, University of Siena, Policlinico S. Maria alle Scotte, Viale Bracci 16, I-53100 Siena, Italy. Tel: +39-0577-233863; Fax: +39-0577-233870; E-mail: zazzi@unisi.it
†Members are listed in the Acknowledgements section.

JAC antiviral

Introduction

Development of resistance to antiretroviral drugs is both a common reason for and a consequence of treatment failure in HIV-1-infected patients. Accordingly, monitoring for the presence of drug resistance mutations in the virus population is recommended in clinical practice both at treatment failure and before initiation of therapy.[1] In most cases, this task is accomplished by using one or more of the many genotype interpretation systems (GISs) freely available over the internet.[2] These typically comprise coded rules or scores whereby any distinct mutation or mutation set is labelled as conferring a defined degree of resistance to one or more individual drugs. Development and maintenance of a GIS is a challenging task because the potentially useful sources of an *in vitro* and *in vivo* evidence base have been growing at an increasing pace and advancements in knowledge of antiretroviral drug resistance must be promptly incorporated into the system.[3] Indeed, the complexity of the correlation between the large number of different mutation sets and the efficacy of the many available treatment regimens are the reason for both the need for such expert systems and the challenge of their continual improvement.

Correlation data between HIV-1 genotype and *in vitro* susceptibility to individual drugs have been a relevant foundation for initial development of the GISs. Following increased availability of *in vivo* treatment-related data, most of the systems have aimed at predicting the efficacy of antiretroviral therapy in HIV-1 patients.[4] *In vivo* findings associating HIV-1 genotype with virological outcome, although sometimes derived from small patient populations, have been in fact typically used for adjusting the GISs in order to increase their clinical utility.[5–7] To be used with confidence, a GIS must be periodically validated to ensure its reliability along with algorithm updates and changes in treatment strategies. In this work, a large data set of HIV-1 genotypes coupled with short- and medium-term virological treatment response data was used to test the performance of four recently updated GISs available over the internet as lists of interpretation rules or fully fledged programs.

Patients and methods

Dataset

The Antiretroviral Resistance Cohort Analysis (ARCA; www.hivarca.net) database, a virtually nationwide observational Italian collection of HIV-1 genotype-centred data subject to patients' informed consent, was used according to the established data accession guidelines to extract treatment change episodes (TCEs). Each TCE was defined by a treatment switch in a patient coupled with a baseline HIV-1 genotype and viral load >5000 HIV-1 RNA copies/mL and a follow-up viral load obtained while on the same uninterrupted therapy. Only TCEs derived from adult patients with complete treatment history information and no mention of major adherence issues were collected. In compliance with the recommendation of the Forum for Collaborative HIV Research,[8,9] the baseline HIV-1 information had to be collected not earlier than 12 weeks before the treatment switch, while two follow-up viral loads where considered for short-term (8 weeks, range 4–12) and medium-term (24 weeks, range 16–32) response. Data for short- and medium-term response were collected independently, i.e. there was no requirement for availability of follow-up information at both timepoints. When multiple data points were available for the same regimen in a patient, the baseline data closer to the start of treatment and the follow-up data closer to the centre of the defined time window were selected. In addition, baseline CD4 counts, HIV-1 subtype, patient demographics as well as several indicators of past antiretroviral drug exposure (number of treatment lines and drugs used grouped by class) were included in the data set. To minimize testing the GISs on obsolete therapies, TCEs were removed when treatment included only nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) or fewer than three drugs [counting any ritonavir-boosted protease inhibitor (PI) as a single drug]. To avoid possible biases in predicting the activity of drugs other than NRTIs, non-nucleoside reverse transcriptase inhibitors (NNRTIs) and PIs in the absence of genotypic information, treatments including any second or later use of enfuvirtide, raltegravir or maraviroc were also excluded. Additional restrictions in the generation of the TCE derived directly from the need to include only drugs considered in all the interpretation systems. Namely, treatments were excluded if they contained zalcitabine or delavirdine, full-dose ritonavir or indinavir, or unboosted PIs other than nelfinavir.

Short-term treatment outcome was dichotomized into success and failure based on achievement and failure to achieve an undetectable viral load or at least a 2 log decrease in viral load at 8 (4–12) weeks. Medium-term success and failure were defined as achieving and not achieving an undetectable viral load at week 24 (16–32). Since the data set included viral load measurements obtained by using laboratory assays with different limits of detection (40, 50, 80 and 500 HIV-1 RNA copies/mL), the 500 copy threshold of sensitivity was used as a cut-off between detectable and undetectable viraemia. Thus, the 5000 HIV-1 RNA copies/mL threshold defined as an inclusion criterion for baseline viral load ensured that minimal decreases, e.g. from 1000 to <500 copies/mL, were not misleadingly labelled as success.

Interpretation systems and genotypic sensitivity scores

The GISs considered included the Stanford HIVdb program (version 5.0.0; http://hivdb.stanford.edu/pages/algs/HIVdb.html), the French ANRS rules (version 17; http://www.hivfrenchresistance.org/2008/Algo-2008.pdf), the Rega rules from Leuven University, Belgium (version 7.1.1; http://www.rega.kuleuven.be/cev/fileadmin/algorithms/Rega_HIV1_Rules_v7.1.1.pdf) and the Italian AntiRetroScan (version 2.0; http://www.hivarca.net/includeGenpub/AntiRetroScan.htm) system running within the ARCA database. ANRS and Rega define three levels of susceptibility (susceptible, possible or intermediate resistant, resistant) for a given HIV-1 genotype to each individual drug based on a set of rules. HIVdb and AntiRetroScan provide five categories of susceptibility or activity (susceptible or complete activity, potential low-level resistance or good activity, low-level resistance or partial activity, intermediate-level resistance or scarce activity, high-level resistance or no activity, respectively). HIVdb and ANRS currently do not give any indication for weighting drug potency. Thus, the predicted efficacy of each drug was scored as 1.00, 0.50 and 0.00 for the three ANRS categories (from susceptible to resistant) and as 1.00, 0.75, 0.50, 0.25 and 0.00 for the five HIVdb categories (from susceptible to high-level resistance). In contrast, the latest versions of Rega and AntiRetroScan introduced drug potency correction factors. With Rega, scoring for the three categories remains 1.00, 0.50 and 0.00 for etravirine, the unboosted PIs and the NRTIs, whereas it is changed to 1.00, 0.25 and 0.00 for enfuvirtide and the other NNRTIs and to 1.50, 0.75 and 0.00 for the ritonavir-boosted PIs. With AntiRetroScan, scoring for the five categories is the

same as in HIVdb but the predicted efficacy score is multiplied by 1.2 for the NNRTIs and by 1.6 for the boosted PIs. To test the added value of drug potency weighting, both Rega and AntiRetroScan were additionally evaluated without using weights for drug potency, i.e. maintaining the same standard 3-level and 5-level 1.00–0.00 scoring scale for all drugs. In addition, the Rega and AntiRetroScan drug potency weighting were also tested for ANRS and HIVdb, respectively. Based on the expectation of full activity at first use, any antiretroviral compound belonging to classes other than NRTIs, NNRTIs and PIs was scored as 1.00 for all the GISs. Each combination regimen was then given a genotypic sensitivity score (GSS) based on the sum of the (weighted) scores coded for the individual drugs included in the regimen.

Statistical analysis

HIV-1 RNA values were log-transformed before analysis. Changes in viral load at 8 and 24 weeks of therapy with respect to baseline were analysed by Wilcoxon signed-rank test.

The associations of the GSS and the other baseline variables with treatment outcome were analysed by univariable and multivariable logistic regression. Variables tested included baseline HIV-1 RNA load, HIV-1 subtype, patient age and gender, HIV-1 transmission risk category, CD4 counts, class-specific and cumulative indicators of previous exposure to treatment, type of TCE (PI based versus NNRTI based) and time elapsed from genotype to the follow-up viral load. Receiver operating characteristic (ROC) curves were computed to evaluate the performance of the success classification scheme based on the GSSs. The Cohen's kappa statistic was computed to measure the inter-rater agreement for all the GIS pairs, based on classification of treatment success or failure at the most accurate GSS threshold calculated for each GIS. Analyses were performed by SPSS version 13.0 (SPSS Inc., Chicago, IL, USA).

Results

Data set characteristics

Based on the inclusion criteria, 765 TCEs from 658 patients and 634 TCEs from 573 patients were generated with virological follow-up data at 8 and 24 weeks, respectively. The median calendar year of treatment start was 2004 [interquartile range (IQR) 2002–2005] for both data sets. Table 1 shows the

Table 1. Baseline characteristics and treatment switch type for the 8 week and 24 week TCEs included in the data set

Feature	8 week TCE data set ($n=65$)	24 week TCE data set ($n=634$)
Median (IQR) patient age, years	41 (37–46)	40 (36–46)
Male gender (%)	507 (66.3)	438 (69.1)
Transmission risk category		
intravenous drug users (%)	251 (32.8)	199 (31.4)
homosexual males (%)	151 (19.7)	120 (18.9)
heterosexual subjects (%)	285 (37.3)	244 (38.5)
other/unknown (%)	78 (10.2)	71 (11.2)
Median (IQR) log plasma HIV-1 RNA, copies/mL	4.59 (4.16–5.10)	4.59 (4.16–5.09)
Median (IQR) CD4 cell count (cells/mm^3)	228 (95–400)	226 (107–383)
Median (IQR) number of previously used drug classes	3 (2–3)	3 (2–3)
Median (IQR) number of past treatment lines	6 (3–10)	6 (3–9)
Median (IQR) number of previously used NRTIs	4 (3–6)	4 (3–5)
Median (IQR) number of previously used NNRTIs	1 (0–1)	1 (0–1)
Median (IQR) number of previously used PIs	2 (1–4)	2 (1–3)
Median (IQR) number of NRTI mutations[10]	3 (1–4)	3 (1–4)
Median (IQR) number of NNRTI mutations[10]	1 (0–2)	1 (0–2)
Median (IQR) number of major PI mutations[10]	0 (0–2)	0 (0–2)
Number (%) of cases including unboosted PI (nelfinavir)	95 (12.4)	87 (13.7)
Number (%) of cases including boosted PI	515 (67.3)	424 (66.9)
Number (%) of cases including NNRTI	226 (29.5)	182 (28.7)
Number (%) of cases including enfuvirtide	69 (9.0)	50 (7.9)
Number (%) of cases including raltegravir	3 (0.4)	0 (0.0)

JAC antiviral

baseline characteristics for the two data sets while Figure 1 details the distribution of the individual drugs in the treatment regimens. Virological response was significant in both the 8 week and 24 week follow-up data sets (median viral load changed from 4.59 to 2.72 and from 4.59 to 2.85 log, respectively; both $P<0.0001$). The proportion of successful treatments according to the study definition was 56% at week 8 and 60% at week 24.

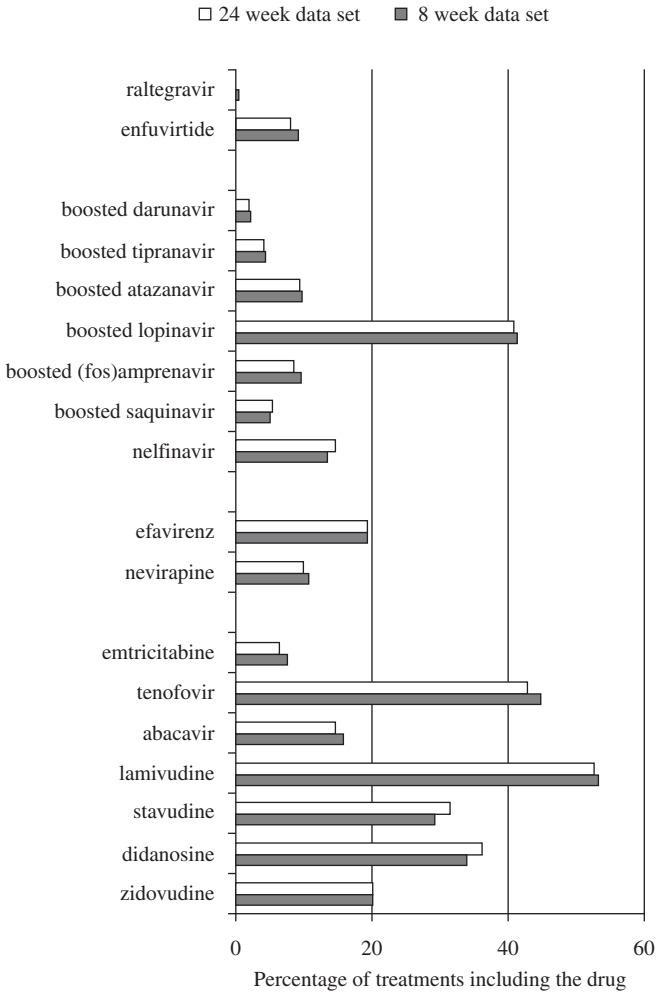

Figure 1. Distribution of NRTIs, NNRTIs, PIs, enfuvirtide and raltegravir in the 8 week and 24 week treatment response data sets.

Prediction of virological outcome by the different GSSs

In the univariable logistic regression analysis, all four GSSs were highly predictive of both 8 week and 24 week virological treatment outcome (Table 2). Higher CD4 counts were associated with increased response at both timepoints while baseline HIV-1 RNA load was associated significantly with 24 week but only marginally with 8 week response. All the indicators of previous drug exposure (number of previous treatment lines, NRTIs, NNRTIs and PIs) were also significantly predictive of worse treatment outcome at both week 8 and 24.

In the multivariable models, each of the GSSs remained highly predictive of treatment outcome at both timepoints (Table 2). A negative association with both 8 week and 24 week treatment outcome in all the analyses was detected for the number of previous PIs used [odds ratio (OR) values 0.57–0.64; $P<0.001$ in all cases]. A higher baseline viral load was predictive of failure in all analyses at week 24 (OR values 0.63–0.70; $P=0.008$ to $P=0.033$) but in none at week 8. Older patient age was predictive of treatment success in all week 8 analyses (OR values 1.27–1.34 per 10 year increase; $P=0.008$ to $P=0.028$) but only with HIVdb and Rega analyses at week 24 (OR 1.26 for both; $P=0.045$ and $P=0.048$, respectively). Undergoing a PI-based treatment switch was associated with success at 24 weeks in the ANRS and HIVdb analyses (OR values 2.65 and 2.50; $P=0.011$ and $P=0.017$, respectively) [see Table S1, available as Supplementary data at *JAC* Online (http://jac.oxfordjournals.org/)]. The accuracies of the GSS-based univariable logistic regression models were 67.3%–69.5% for the 8 week outcome and 63.2%–67.0% for the 24 week outcome. In the multivariable models, the accuracy increased to 71.7%–73.7% for the 8 week outcome and to 73.5%–75.1% for the 24 week outcome.

The ROC curves in Figure 2 show the tradeoff between sensitivity and specificity for each GSS as a predictor of the treatment success. Based on the area under the ROC curve (AUROC) value, there were some differences in performance. AntiRetroScan was the most accurate of the systems with both the 8 week ($P=0.001$ versus ANRS; $P=0.016$ versus Rega;

Table 2. Crude and adjusted ORs for virological success depending on the different GSSs

	8 week data set[a]		24 week data set[a]	
GSS	crude OR (95% CI)	adjusted OR[b] (95% CI)	crude OR (95% CI)	adjusted OR[c] (95% CI)
HIVdb	2.39 (2.01–2.85)	2.21 (1.80–2.71)	2.21 (1.83–2.67)	2.14 (1.70–2.70)
AntiRetroScan	2.49 (2.10–2.95)	2.26 (1.87–2.74)	2.52 (2.09–3.06)	2.43 (1.94–3.05)
ANRS	2.31 (1.93–2.75)	2.18 (1.78–2.67)	2.25 (1.85–2.73)	2.21 (1.76–2.78)
Rega	2.29 (1.93–2.71)	2.04 (1.69–2.46)	2.28 (1.88–2.76)	2.15 (1.72–2.69)

OR values are per unit increase of each GSS.
[a]All P values <0.0001.
[b]Additional significant predictors of success include older age (all GISs). Additional significant predictors of failure include a larger number of previously used PIs (all GISs). See Table S1 for details.
[c]Additional significant predictors of success include older age (HIVdb, Rega) and undergoing a PI-based treatment switch (ANRS, HIVdb). Additional significant predictors of failure include a higher baseline viral load (all GISs) and a larger number of previously used PIs (all GISs). See Table S1 for details.

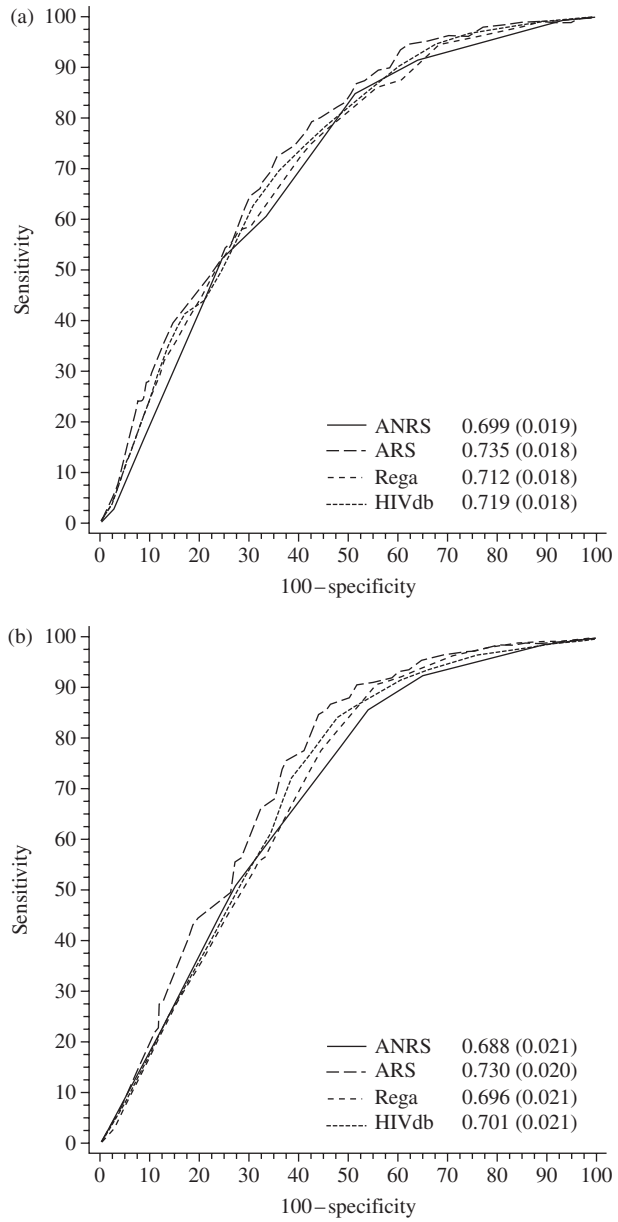

Figure 2. ROC curves for the four GSSs as predictors of treatment success at 8 (a) and 24 (b) weeks. The inserts show the values of the area under the curve with standard errors in parentheses. ARS, AntiRetroScan.

$P=0.045$ versus HIVdb) and 24 week ($P<0.001$ versus ANRS; $P=0.003$ versus HIVdb; $P=0.004$ versus Rega) data set. The only other significant difference was HIVdb outperforming ANRS ($P=0.048$) in the 8 week data set. The interval where AntiRetroScan outperformed the other GISs was at 55%–95% sensitivity corresponding to 30%–75% specificity with the 8 week outcome and at 25%–90% sensitivity corresponding to 50%–90% specificity with the 24 week outcome. Table 3 shows the sensitivity and specificity of the prediction of 8 week and 24 week success by each GIS at the GSS threshold of 3.0, which is typically recommended when building a highly effective antiretroviral regimen. Sensitivity and specificity are comparatively indicated at the GSS threshold corresponding to the highest accuracy of each GIS. Interestingly, the most accurate GSS threshold for all the GISs and at both timepoints was <3.0.

To test whether the response to NNRTI- and PI-based treatments was predicted with different accuracy by the GISs under study, ROC curves were then computed after stratifying for treatment type. The few cases (8.8%) with treatment comprising both a PI and an NNRTI were included in both strata. As shown in Table 4, the apparent overall superiority of AntiRetroScan could be attributed to a significantly better performance with PI-based treatments ($P=0.003$ and $P=0.003$ versus ANRS, $P=0.105$ and $P=0.003$ versus HIVdb, $P=0.042$ and $P=0.005$ versus Rega with the 8 week and 24 week data set, respectively). In contrast, there were no differences in accuracy between any GIS pair with NNRTI-based treatments.

Agreement between the GISs in the classification of treatment failure and success

The TCEs were classified as success or failure by each GIS using its own most accurate GSS value as a cut-off (Table 4). The kappa values for the agreement between the GISs in the classification of treatment failure and success in the 8 week data set were 0.608 for ANRS and AntiRetroScan, 0.611 for ANRS and HIVdb, 0.521 for ANRS and Rega, 0.749 for AntiRetroScan and HIVdb, 0.684 for AntiRetroScan and Rega, and 0.728 for HIVdb and Rega. Thus, ANRS was apparently the more divergent system with this short-term response data set. However, the difference was no longer detected with the 24 week data set (kappa values 0.681 for ANRS and AntiRetroScan, 0.737 for ANRS and HIVdb, 0.655 for ANRS and Rega, 0.725 for

Table 3. Sensitivity and specificity of the prediction of treatment success at week 8 and 24 by each GIS at the GSS threshold of 3.0 and at the GSS threshold corresponding to maximum accuracy

	8 week response					24 week response				
	GSS threshold = 3.0		most accurate GSS threshold			GSS threshold = 3.0		most accurate GSS threshold		
GIS	sensitivity	specificity	GSS	sensitivity	specificity	sensitivity	specificity	GSS	sensitivity	specificity
HIVdb	36.83	85.42	1.50	70.4	63.39	33.88	81.35	1.25	84.36	51.99
AntiRetroScan	47.55	78.57	2.20	72.49	63.49	47.56	76.45	1.95	86.97	53.82
ANRS	52.45	75.30	1.50	85.08	48.21	51.47	72.48	1.50	85.67	45.87
Rega	58.04	71.73	2.25	73.66	58.93	55.70	68.20	1.75	90.88	44.34

JAC antiviral

Table 4. AUROC for the different GSSs with respect to 8 week and 24 week response to NNRTI- and PI-based treatment

GSS	8 week AUROC (standard error)		24 week AUROC (standard error)	
	NNRTI based	PI based	NNRTI based	PI based
ANRS	0.730 (0.033)	0.692 (0.021)	0.722 (0.038)	0.684 (0.023)
AntiRetroScan	0.732 (0.033)	0.728 (0.020)	0.744 (0.037)	0.724 (0.022)
HIVdb	0.737 (0.033)	0.713 (0.021)	0.753 (0.037)	0.695 (0.023)
Rega	0.727 (0.033)	0.705 (0.021)	0.730 (0.038)	0.687 (0.023)

AntiRetroScan and HIVdb, 0.698 for AntiRetroScan and Rega, and 0.706 for HIVdb and Rega).

Taking into account the real treatment outcome, the four GISs correctly and unanimously predicted 70.6% of the successful treatments and 32.4% of the unsuccessful treatments at 8 weeks, and 62.5% of the successful treatments and 26.4% of the unsuccessful treatments at 24 weeks. Incorrect prediction by all the systems was far more common with treatment failure (33.0% at 8 weeks, 45.2% at 24 weeks) than with treatment success (7.5% at 8 weeks, 4.7% at 24 weeks) [see Table S2, available as Supplementary data at *JAC* Online (http://jac.oxfordjournals.org/)].

Impact of drug potency weighting

Since both AntiRetroScan and Rega implement drug potency weighting factors, the two systems were each then compared with its own version devoid of the correction factors in order to test whether this strategy actually improves the predictive power. Comparison of ROC curves revealed that both GISs improve significantly with the adoption of these drug potency weights (Figure 3). With the 8 week data set, the AUROC values increased from 0.717 to 0.735 for AntiRetroScan and from 0.687 to 0.712 for Rega ($P=0.001$ for AntiRetroScan; $P<0.001$ for Rega). With the 24 week data set, the AUROC values increased from 0.703 to 0.730 for AntiRetroScan and from 0.663 to 0.696 for Rega (both $P<0.001$). The larger AUROC with the unweighted version of AntiRetroScan with respect to the unweighted version of Rega was again due to improved performance with the PI-based treatments (data not shown).

Finally, although the current versions of HIVdb and ANRS do not use any drug class correction factor to compute the GSS, the potential benefit from adopting a drug potency weighting approach was also investigated for these systems. Due to the different number of drug susceptibility levels used (three for ANRS and Rega, five for HIVdb and AntiRetroScan), ANRS was tested with the Rega weights and HIVdb was tested with the AntiRetroScan weights. The AUROC of the ANRS algorithm increased significantly when the GSS was adjusted according to the Rega drug weighting (from 0.699 to 0.730 at 8 weeks, $P<0.001$; from 0.688 to 0.721 at 24 weeks, $P<0.001$). The AUROC of the HIVdb algorithm increased significantly when the GSS was adjusted according to the AntiRetroScan drug weighting (from 0.719 to 0.739 at 8 weeks, $P<0.001$; from 0.701 to 0.726 at 24 weeks, $P<0.001$).

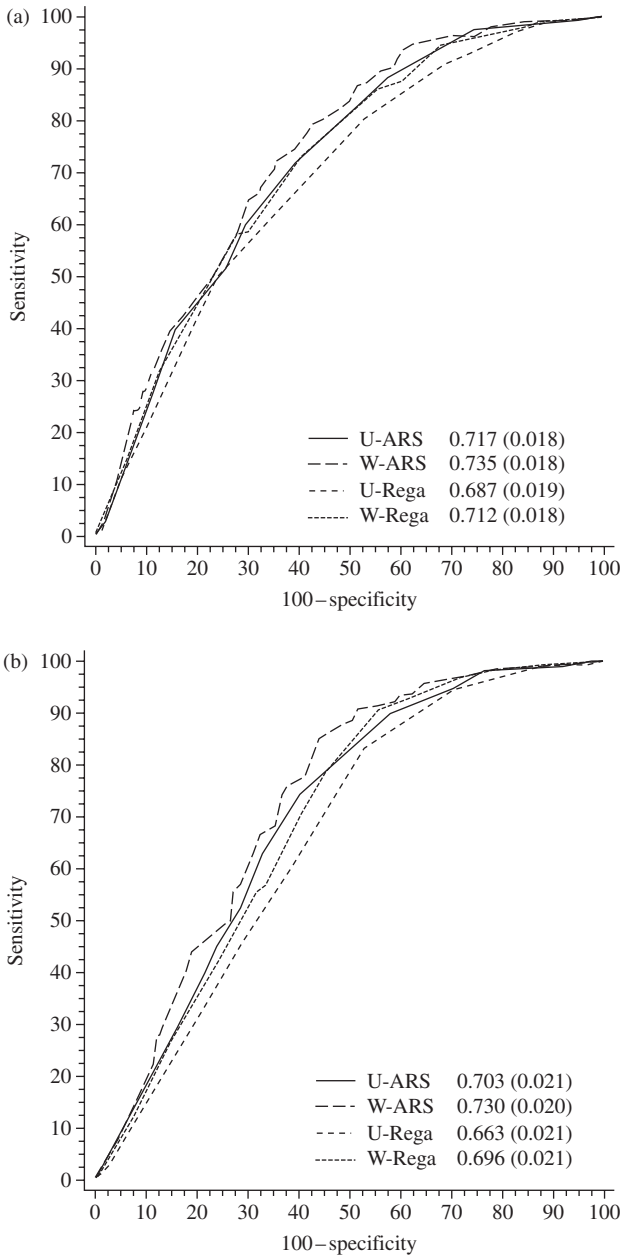

Figure 3. ROC curves for the drug potency unweighted (U) and weighted (W) Rega and AntiRetroScan (ARS) GSSs as predictors of treatment success at 8 (a) and 24 (b) weeks. The inserts show the values of the area under the curve with standard errors in parentheses.

Discussion

Genotypic antiretroviral resistance is widely recognized as a major factor impacting response to antiretroviral therapy. Accordingly, a number of tools have been developed to interpret HIV-1 mutational patterns and assist the choice of antiretroviral regimens.[2] Comparisons of a large number of GISs have been previously reported.[4,8,9] In this work, the updated versions of the three most popular systems plus the ARCA built-in algorithm[11] were assessed for their ability to predict virological outcome of a large set of TCEs derived from clinical practice in Italy. Two follow-up observations were considered, in compliance with the Forum for Collaborative HIV Research recommendation for short-term and medium-term response.[8] Definition of virological treatment success can vary in different studies. The definition adopted in this study was based on the reasonable expectation that an effective treatment allows control of viral replication to undetectable levels. Due to inclusion of viral load measurements obtained with first-generation assays, the detection threshold used was 500 HIV-1 RNA copies/mL. However, among the cases with undetectable follow-up viral load, only 18.6% at week 8 and 12.2% at week 24 were obtained with the 500 copy threshold. Indeed, comparable results were obtained when the analysis was repeated with the current definition of success at <50 HIV-1 RNA copies/mL or at any undetectable viral load (data not shown). Two design adjustments were used to minimize mislabelling of successful cases. First, TCEs with baseline viral load <5000 copies/mL were excluded, avoiding scoring as success a limited decrease in viral load. Secondly, a decrease of at least 2 log in viral load at week 8 (4–12) was scored as a success, allowing the capture of an effective therapy not achieving complete control of viral replication in the short term because of high-level baseline viral load. Finally, virological outcomes of first-line therapy on drug-naïve patients were excluded in order to build a more challenging data set. Analysis of a 30% enlarged data set also comprising first-line therapy TCEs indeed resulted in a 5–7% increase in the AUROC for prediction of 24 week outcome (data not shown).

All of the GSSs were significantly associated with virological treatment outcome at both timepoints, according to logistic regression analysis and ROC curve analysis. Multivariable logistic regression indicated that each of the GSSs remained a highly significant predictor after adjusting for patient demographics, baseline CD4 count, viral load, HIV-1 subtype and several indicators of previous treatment exposure. This confirms that HIV-1 genotype, as interpreted by currently available rules-based systems, independently impacts virological outcome, despite the simplistic summation of the individual drug scores into a summary GSS. Although different TCE definition and treatment selection criteria do not allow direct comparisons, current ANRS, HIVdb and Rega GSSs appear to perform better than their own earlier versions used in a previous similar study.[4] However, confirmation of an increased accuracy of the GISs over time clearly requires running older and newer algorithm versions on the same data set. In addition, it will be advisable to test the GISs for their ability to predict treatment response at longer follow-up times, such as 48 or 96 weeks. Currently, historical data sets include a large proportion of therapies with a shorter duration but, since modern antiretroviral regimens are more convenient and generally better tolerated than the older ones, it will be probably possible to evaluate the performance of the different GISs with more clinically relevant long-term response.

ROC curve analysis allowed a direct comparison of the systems and revealed that AntiRetroScan was more accurate than the other systems in predicting PI-based treatment outcome. Rega and AntiRetroScan have recently introduced drug or drug class weighting factors in an attempt to correct the GSS for the relative potency of available antiretrovirals. Although the currently used correction factors have been arbitrarily set based on expert view rather than derived from statistical learning, it is interesting to note that both the GISs improved significantly with respect to the corresponding unweighted version. In addition, when the Rega and AntiRetroScan drug potency weights were applied to the ANRS and HIVdb systems, respectively, both of these also improved significantly. This provides the basis for further work along these lines. Post-hoc analysis with the 24 week data set (not shown) indicated that the AUROC for AntiRetroScan continued to increase when using up to 2.2 as a potency weighting factor for boosted PIs. Conversely, no improvement was obtained by increasing the NNRTI potency weighting factor. This is in line with recent data and systematic reviews showing an advantage with boosted PI- versus NNRTI-based therapy, particularly in the setting of drug resistance.[12,13] Further confirming the major role of boosted PIs in this setting, the extent of previous exposure to PIs remained a strong independent predictor of both 8 week and 24 week treatment failure together with the GSS in all the multivariable logistic regression analyses. In addition, undergoing a PI-based treatment change was predictive of virological success at week 24 in the analyses of two GISs.

The case files used for this study were derived from clinical practice, and one or more of the GISs considered may have been preferentially used to build treatment regimens. Thus, records might be an over-represented collection of expected successes according to a specific GIS(s). In particular, almost 30% of TCEs were derived from clinics receiving the AntiRetroScan report from a single reference laboratory as a response to the genotypic test requested. However, ROC curve analysis of the TCE subset excluding these records confirmed the results obtained on the whole data set (data not shown). Although AntiRetroScan is associated with the ARCA initiative as a built-in service for the units uploading data to the web server, no use of these data and thus of any derived TCE-like information has been done for training of AntiRetroScan. The system in fact uses a purely rules-based approach which relies on periodical analysis of HIV-1 drug resistance-related literature and is not derived from any statistical learning procedure. Therefore, the TCE data sets used in this study were equally unknown for all the GISs examined and cannot have contributed to the increased accuracy of AntiRetroScan.

The data sets have several limitations. First, obsolete therapies were largely excluded but novel drugs such as darunavir and etravirine were under- or not represented. Secondly, drugs belonging to classes other than NRTIs, NNRTIs and PIs were considered active by definition in the absence of genotypic information since they were being used for the first time. While this appears a reasonable expectation in principle, it cannot be excluded that these drugs did not work properly in occasional patients. Thirdly, information on patient adherence from the ARCA centres was not consistently available. The database does not store the method for definition of adherence. The patient

JAC antiviral

records marked with adherence issues were excluded, but evaluation of adherence may have been different in different centres and probably not all adherence issues were originally indicated. Inclusion of TCEs derived from non-adherent patients may thus have generated treatment failures unrelated to drug resistance and decreased the specificity of prediction of success. Indeed, incorrect predictions by all four systems were far more common with treatment failure than with treatment success.

Although all the GISs were confirmed to be useful tools for guiding treatment choices, it must be noted that in absolute terms their accuracy remains limited, at least in pre-treated patients and with the success definitions used here. For example, predicting treatment success with 80% sensitivity occurs at only 55%–60% specificity. It is presently uncertain whether significant improvements in rules-based systems can be obtained and allow more reliable prediction of treatment outcome based on HIV-1 genotype only. Indeed, the accuracy of the virological outcome prediction increased significantly in the multivariable with respect to the GSS-based univariable models. Accordingly, ongoing studies with large databases suggest that including additional patient and virus features as input variables can increase accuracy through several modelling techniques.[14–17] While such prototype systems are expected to mature into clinical tools, commonly used and regularly updated rules-based GISs remain a cornerstone in assisting antiretroviral treatment choices and appear to be amenable to further improvement with the inclusion of drug potency weighting factors.

Acknowledgements

This work has been presented at the Seventh European Workshop on HIV Drug Resistance, Stockholm, 2009.

The following investigators take part in the ARCA database: Andrea Giacometti (Ancona - Clinica di Malattie Infettive), Luca Butini (Ancona - Immunologia Clinica), Romana del Gobbo (Ancona - Malattie Infettive), Stefano Menzo (Ancona - Virologia), Danilo Tacconi (Arezzo - Malattie Infettive), Giovanni Corbelli (Ascoli Piceno - Malattie Infettive), Stefania Zanussi (Aviano - Centro di Riferimento Oncologico), Laura Monno (Bari - Clinica Malattie Infettive Università), Grazia Punzi (Bari - Virologia), Franco Maggiolo (Bergamo - Malattie Infettive), Annapaola Callegaro (Bergamo - Microbiologia e Virologia), Leonardo Calza (Bologna - Malattie Infettive S. Orsola), Maria Carla Re (Bologna - UO Microbiologia, Lab. Retrovirus), Raffaele Pristerà (Bolzano - Malattie Infettive), Paola Turconi (Brescia - Fleming Labs), Antonella Mandas (Cagliari - Centro S.I.D.A., Policlinico Universitario), Sauro Tini (Citta' di Castello - Medicina Generale), Giuseppe Carnevale (Cremona - Malattie Infettive), Elisabetta Paolini (Cremona - Servizio Immunoematologia e Medcina Trasfusionale), Giorgio Amadio (Fermo - Malattie Infettive), Laura Sighinolfi (Ferrara - Malattie Infettive AOU S. Anna), Giuliano Zuccati (Firenze - Centro MTS), Massimo Morfini (Firenze - Ematologia Careggi), Roberto Manetti (Firenze - Immunoallergologia Careggi), Paola Corsi (Firenze - Malattie Infettive Careggi), Luisa Galli (Firenze - Malattie Infettive Pediatria Meyer), Massimo Di Pietro (Firenze - Malattie Infettive SM Annunziata), Filippo Bartalesi (Firenze - Malattie Infettive Università), Grazia Colao (Firenze - Virologia Careggi), Andrea Tosti (Foligno - Malattie Infettive/SERT), Antonio Di Biagio (Genova - Clinica Malattie Infettive AOU S. Martino), Maurizio Setti (Genova - Clinica Medica Immunologia), Bianca Bruzzone (Genova - Laboratorio di Igiene Ospedale S. Martino), Giovanni Penco (Genova - Malattie Infettive Ospedali Galliera), Michele Trezzi (Grosseto - Malattie Infettive), Anna Orani (Lecco - Malattie Infettive), Riccardo Pardelli (Livorno - Malattie Infettive), Michele De Gennaro (Lucca - Malattie Infettive), Alessandro Chiodera (Macerata - Malattie Infettive), Alfredo Scalzini (Mantova - Malattie Infettive Ospedale 'C. Poma'), Loredana Palvarini (Mantova - Virologia), Paolo Almi (Massa - Malattie Infettive), Giovanni Todaro (Messina - Malattie Infettive), Antonella d'Arminio Monforte (Milano - Clinica Malattie Infettive AO S. Paolo), Paola Cicconi (Milano - Clinica di Malattie Infettive Ospedale S. Paolo), Stefano Rusconi (Milano - Dipart. Scienze Cliniche, Sez. Malattie Infettive, Università degli Studi), Maria Rita Gismondo [Milano - Laboratorio Microbiologia Ospedale L. Sacco (Dipart. Scienze Cliniche, Sez. Malattie Infettive)], Maria Rita Gismondo [Milano - Laboratorio Microbiologia Ospedale L. Sacco (Prima Divisione Malattie Infettive)], Valeria Micheli [Milano - Laboratorio Microbiologia Ospedale L. Sacco (Seconda Divisione Malattie Infettive)], Maria Luisa Biondi (Milano - Laboratorio di diagnostica molecolare infettivologica AO S. Paolo), Nicola Gianotti (Milano - Malattie Infettive San Raffaele), Amedeo Capetti (Milano - Prima Divisione Malattie Infettive Ospedale L. Sacco), Paola Meraviglia (Milano - Seconda Divisione Malattie Infettive Ospedale L. Sacco), Enzo Boeri (Milano - Virologia HSR), Cristina Mussini (Modena - Clinica Malattie Infettive), Monica Pecorari (Modena - Virologia), Maurizio Santirocchi (Narni - SERT), Diego Brustia (Novara - Malattie Infettive AO Maggiore), Paolo Ravanini (Novara - Virologia), Federico Dal Bello (Padova - Virologia), Nino Romano (Palermo - Centro Riferimento AIDS Università), Salvatrice Mancuso (Palermo - Servizio Riferimento Regionale Diagnosi AIDS), Carlo Calzetti (Parma - Divisione Malattie Infettive ed Epatologia Azienda Ospedaliera), Renato Maserati (Pavia - Ambulatorio Clinica Malattie Infettive S. Matteo), Gaetano Filice (Pavia - Clinica Malattie Infettive e Tropicali), Fausto Baldanti (Pavia - Virologia S. Matteo), Daniela Francisci (Perugia - Malattie Infettive), Giustino Parruti (Pescara - Malattie Infettive), Ennio Polilli (Pescara - Virologia Pescara), Daria Sacchini (Piacenza - Malattie Infettive), Chiara Martinelli (Pisa - Malattie Infettive), Rita Consolini (Pisa - Pediatria I Università), Linda Vatteroni (Pisa - Virologia), Angela Vivarelli (Pistoia - Malattie Infettive), Daniele Dionisio (Pistoia - Virologia), Alessandro Nerli (Prato - Malattie Infettive), Lucia Lenzi (Prato - Virologia), Giacomo Magnani (Reggio Emilia - Malattie Infettive), Patrizia Ortolani (Rimini - Malattie Infettive Rimini), Massimo Andreoni (Roma - Cattedra Malattie Infettive Tor Vergata), Guido Palamara (Roma - IRCCS S. Gallicano), Caterina Fimiani (Roma - Immunologia Clinica Umberto I), Lucia Palmisano (Roma - Istituto Superiore di Sanità), Andrea De Luca (Roma - Istituto di Clinica Malattie Infettive Cattolica), Simona Di Giambenedetto (Roma - Laboratorio virologia Cattolica), Andrea Antinori (Roma - Malattie Infettive INMI Spallanzani), Vincenzo Vullo (Roma - Malattie Infettive e Tropicali La Sapienza - Umberto I), Ombretta Turriziani (Roma - Medicina Sperimentale e Patologia - Sezione Virologia - La Sapienza), Carlo Federico Perno (Roma - Monitoraggio Terapie Antivirali e Antineoplastiche INMI Spallanzani), Carlo Federico Perno (Roma - Virologia DMS Tor Vergata), Marco Montano (Roma - Virologia per Malattie

Infettive Tor Vergata), Giovanni Cenderello (San Remo - Malattie Infettive), Angela Gonnelli (Siena - Malattie Infettive), Laura Romano (Siena - Virologia), Michele Palumbo (Terni - Malattie Infettive), Valeria Ghisetti (Torino - Laboratorio di Virologia, Ospedale Amedeo di Savoia), Stefano Bonora (Torino - Malattie Infettive Amedeo di Savoia), Palma Delle Foglie (Trento - Malattie Infettive), Cristina Rossi (Treviso - Malattie Infettive), Federica Poletti (Verbania - Malattie Infettive Verbania), Vincenzo Mondino (Verbania - Virologia), Marina Malena (Verona - Centro di Medicina Preventiva-ULSS 20), Emanuela Lattuada (Verona - Malattie Infettive).

Funding

This work was supported by the VI AIDS Research Program of the Ministry of Health, Rome, Italy (grant 30G.58). ARCA is supported by educational grants from Abbott, Boehringer-Ingelheim, Bristol-Myers Squibb, GlaxoSmithKline, Gilead Sciences, Janssen-Cilag Tibotec division.

Transparency declarations

A. D. L. has been a member of advisory boards or has received speaker fees from GlaxoSmithKline, Bristol-Myers Squibb, Gilead Sciences, Abbott, Boehringer Ingelheim, Tibotec, Monogram Biosciences and Siemens Diagnostics. He is also a member of the ARCA Scientific Board.

F. M. has served as a consultant on advisory boards for Boehringer Ingelheim, Bristol-Myers Squibb, Gilead, GlaxoSmithKline, Roche and Tibotec, has received lecture fees from Abbott, Bayer, Bristol-Myers Squibb, Gilead, GlaxoSmithKline, Merck Sharp and Dohme, Pfizer and Roche, and has received research and educational grants from Boehringer Ingelheim, Bristol-Myers Squibb, GlaxoSmithKline, Janssen-Cilag and Roche.

M. P. is a member of the ARCA Scientific Board.

S. R. has received research support or honoraria from Abbott, Boehringer Ingelheim, Bristol-Myers Squibb, Gilead, GlaxoSmithKline and Pfizer, has served as a consultant to Abbott, Janssen-Cilag and Boehringer Ingelheim, and is also a member of the ARCA Scientific Board.

M. Z. has been a consultant to Abbott Molecular, Gilead Sciences and Boehringer Ingelheim, has received research support or lecture fees from Abbott Pharmaceuticals, Merck Sharp and Dohme, Pfizer and Janssen-Cilag, and is also a member of the ARCA Scientific Board and the creator of the AntiRetroScan system.

All other authors: nothing to declare.

Supplementary data

Tables S1 and S2 are available as Supplementary data at *JAC* Online (http://jac.oxfordjournals.org/).

References

1. Hammer SM, Eron JJ Jr, Reiss P *et al*. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. *JAMA* 2008; **300**: 555–70.

2. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis* 2006; **42**: 1608–18.

3. Van Laethem K, Vandamme AM. Interpreting resistance data for HIV-1 therapy management—know the limitations. *AIDS Rev* 2006; **8**: 37–43.

4. 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; **187**: 1934–43.

5. King MS, Rode R, Cohen-Codar I *et al*. Predictive genotypic algorithm for virologic response to lopinavir-ritonavir in protease inhibitor-experienced patients. *Antimicrob Agents Chemother* 2007; **51**: 3067–74.

6. De Luca A, Giambenedetto SD, Trotta MP *et al*. Improved interpretation of genotypic changes in the HIV-1 reverse transcriptase coding region that determine the virological response to didanosine. *J Infect Dis* 2007; **196**: 1645–53.

7. Marcelin AG, Flandre P, Molina JM *et al*. Genotypic resistance analysis of the virological response to fosamprenavir-ritonavir in protease inhibitor-experienced patients in CONTEXT and TRIAD clinical trials. *Antimicrob Agents Chemother* 2008; **52**: 4251–7.

8. Cozzi-Lepri A. Initiatives for developing and comparing genotype interpretation systems: external validation of existing rule-based interpretation systems for abacavir against virological response. *HIV Med* 2008; **9**: 27–40.

9. Assoumou L, Brun-Vézinet F, Cozzi-Lepri A *et al*. Initiatives for developing and comparing genotype interpretation systems: external validation of existing systems for didanosine against virological response. *J Infect Dis* 2008; **198**: 470–80.

10. Johnson VA, Brun-Vezinet F, Clotet B *et al*. Update of the drug resistance mutations in HIV-1: spring 2008. *Top HIV Med* 2008; **16**: 62–8.

11. Zazzi M, Romano L, Venturi G *et al*. Comparative evaluation of three computerized algorithms for prediction of antiretroviral susceptibility from HIV type 1 genotype. *J Antimicrob Chemother* 2004; **53**: 356–60.

12. Shen L, Peterson S, Sedaghat AR, McMahon MA *et al*. Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. *Nat Med* 2008; **14**: 762–6.

13. Gupta R, Hill A, Sawyer AW *et al*. Emergence of drug resistance in HIV type 1-infected patients after receipt of first-line highly active antiretroviral therapy: a systematic review of clinical trials. *Clin Infect Dis* 2008; **47**: 712–22.

14. Lengauer T, Sing T. Bioinformatics-assisted anti-HIV therapy. *Nat Rev Microbiol* 2006; **4**: 790–7.

15. Altmann A, Beerenwinkel N, Sing T *et al*. Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. *Antivir Ther* 2007; **12**: 169–78.

16. Larder B, 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.

17. Rosen-Zvi M, Altmann A, Prosperi M *et al*. Selecting anti-HIV therapies based on a variety of genomic and clinical factors. *Bioinformatics* 2008; **24**: i399–406.