Computational models developed without a genotype for resource-poor countries predict response to HIV treatment with 82% accuracy

AD Revell, D Wang, R Harrigan, J Gatell, L Ruiz, S Emery, MJ Pérez-Elías, C Torti, J Baxter, F DeWolf, B Gazzard[1], AM Geretti, S Staszewski, R Hamers, AMJ Wensing, J Lange, JM Montaner, BA Larder

HIV Resistance Response Database Initiative

The clinical issue

- Combination antiretroviral therapy (cART) is being rolled out in resource-poor countries
- Treatments are failing at a comparable rate to other countries with resistance a significant factor
- Selecting the optimum drug combination after failure in these settings is a major challenge:
  - Resistance testing is not widely available
  - Treatment options are limited
  - Healthcare provider experience may be limited
- Could the RDI's approach be of help?

What is the RDI's approach?

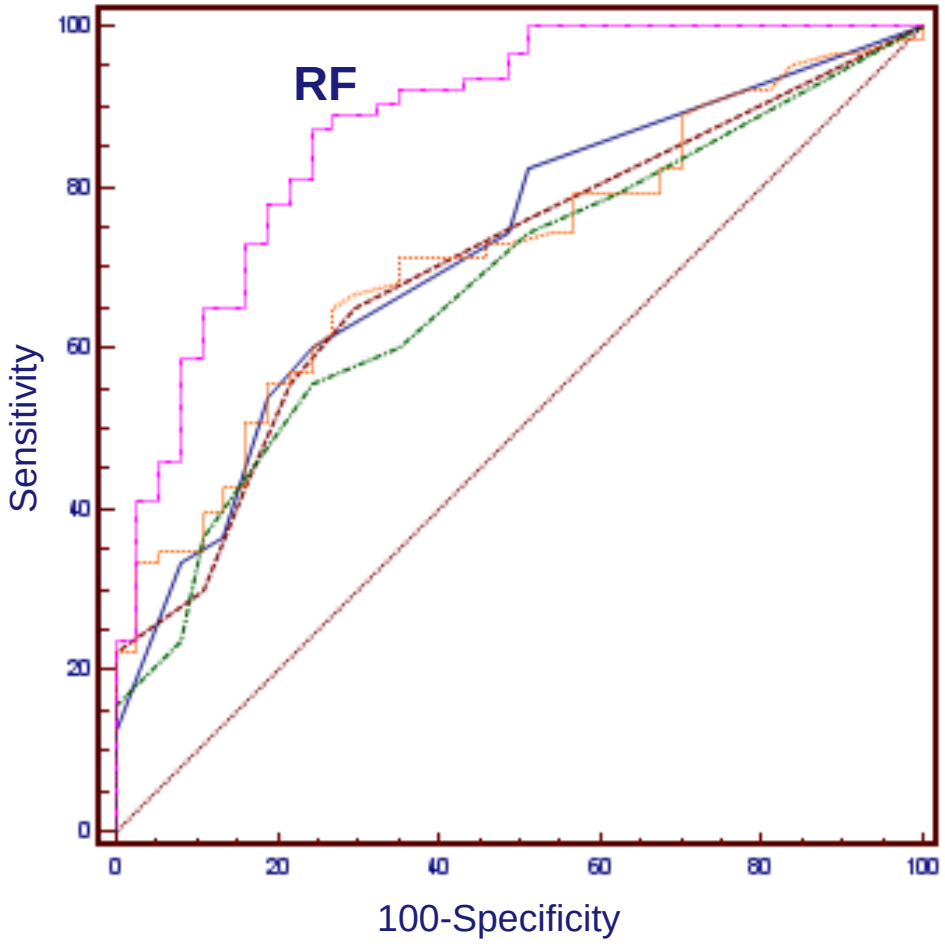
To develop computational models using data from many '000s patients to predict response to cART:

- **Initially: the change in viral load from baseline following a treatment change**
  - Correlation of predicted vs actual virological response typically gave $r^2 \geq 0.70$ and mean difference of <0.5 log copies/ml
- **Recently: the probability that the viral load will go 'undetectable' (<50 copies/ml)**

RF model developed to predict probability of VL<50 copies

- 3,188 training treatment change episodes (TCEs) & 100 test TCEs used
- The RDI's 'standard' set of 82 input variables, including 58 mutations plus BL VL, CD4, treatment history, drugs in new regimen and time to follow-up
- Predictive accuracy compared with performance of genotypic sensitivity scores (GSS) derived from current 'rules' systems for interpretation of genotype

ROC curve for RF model & GSS from common rules systems predicting VL<50 copies



RDI RF: AUC = 0.88
Accuracy = 82%

GSS: AUC = 0.68-0.72
Accuracy = 63-68%

Current study objectives

1. To develop RF models to predict virological response to cART (VL<50 copies) without the use of genotype
2. To use a large dataset representative of clinical practice in resource-poor countries
3. To use the models to identify potentially effective alternative regimens for cases of actual virological failure

Data selection/partition

- 8,514 TCEs from > 20 centres in ‘rich countries’ selected from RDI database
- No historical exposure to PIs, T-20, raltegravir or maraviroc but PIs allowed in the new regimen (to represent typical clinical practice in resource-poor countries)
- Data partitioned at random by patient into 8,114 training and 400 test TCEs

Datasets - descriptive statistics

	Training (8,114 TCEs)	Test (400 TCEs)
Male	6,410 (79%)	308 (77%)
Median BL viral load (log copies HIV RNA/ml)	2.27	3.04
Median BL CD4 (cells/ml)	316	287
Number of different regimens (new treatment)	248	51
Number (percent) failures	3,215 (40%)	205 (51%)

Developing the models

Two RF models were trained to predict the probability of the follow-up viral load being <50 copies:

Model 1

24 Input variables:

- Baseline viral load
- Baseline CD4 count
- Treatment history (AZT, 3TC, any NNRTI)
- Drugs in the new regimen
- Time to follow-up

Model 2

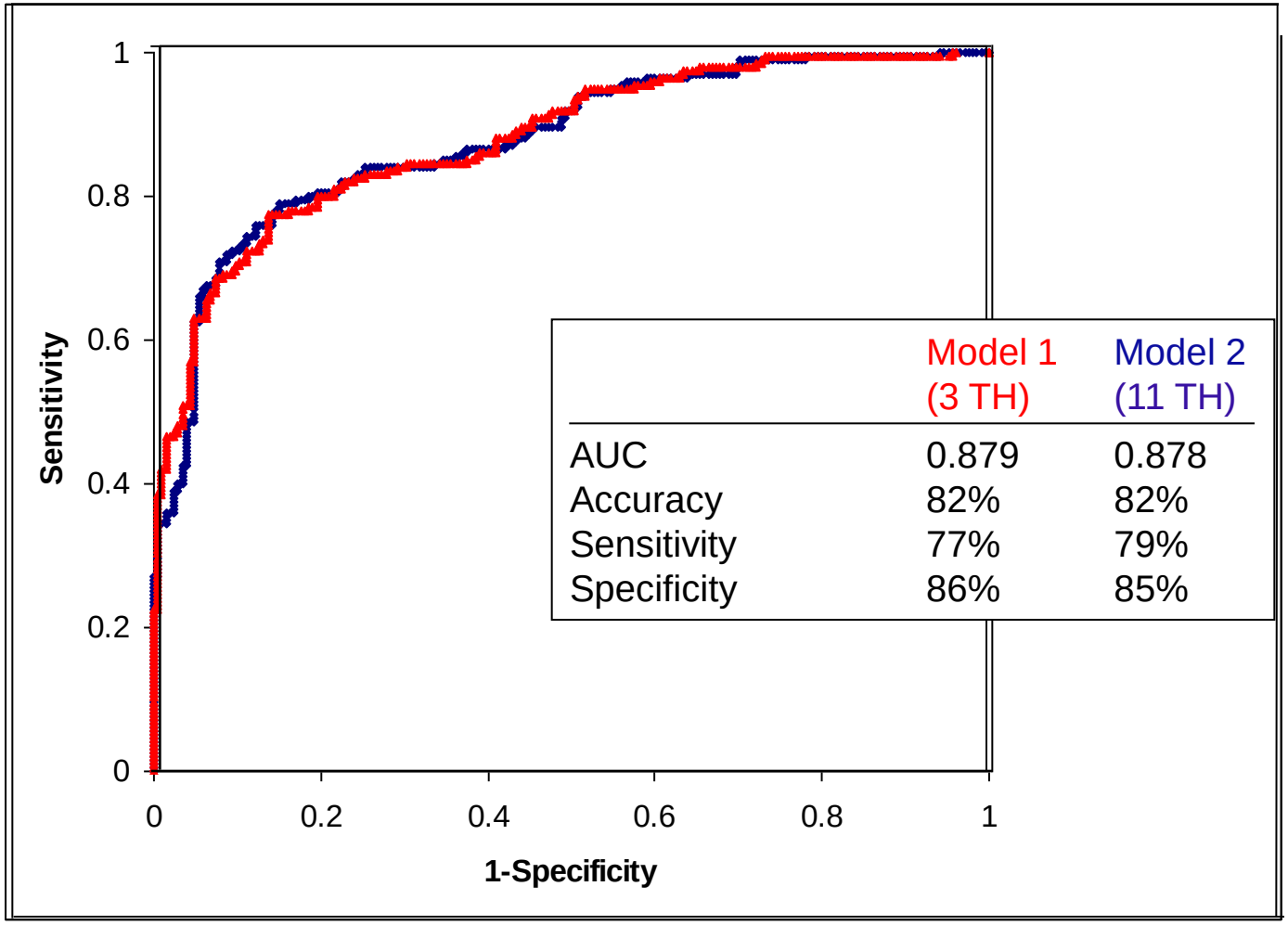
32 Input variables:

- As Model 1 except 11 individual drug treatment history variables were used.

Testing the models

- RF models analysed baseline data from test TCEs
- Produced estimate of probability of the follow-up VL being <50 copies
- ROC curves plotted for models' predictions *vs* actual responses

ROC curve



	Model 1 (3 TH)	Model 2 (11 TH)
AUC	0.879	0.878
Accuracy	82%	82%
Sensitivity	77%	79%
Specificity	86%	85%

Relative importance of input variables for modelling virological response (Model 2)

Input variable	Importance rank	Importance score
Baseline viral load	1	150.67
Time to follow-up	2	42.58
Baseline CD4 count	3	33.56
EFV - historical	4	29.98
TDF - current	5	25.17
ddI - current	6	20.04
AZT - historical	7	18.32
d4T - current	8	18.20
AZT - current	9	17.87
3TC - historical	10	17.45

In silico analysis

- Models were programmed to predict responses to multiple alternative 3-drug regimens using baseline data from the cases where the new treatment failed using two drug lists:
  - ~'Old' PIs only (IDV/r, SQV/r, LPV/r, NFV)
  - ~Including 'newer' PIs ((fos-)APV/r, ATZ/r, DRV/r)

In silico analysis

- Models were programmed to predict responses to multiple alternative 3-drug regimens using baseline data from the cases where the new treatment failed using two drug lists:
  - ~'Old' PIs only (IDV/r, SQV/r, LPV/r, NFV)
  - ~Including 'newer' PIs ((fos-)APV/r, ATZ/r, DRV/r)

	Model 1 (3 TH)	**Model 2 (11 TH)**
% failures for which effective alternatives found (old drugs only)	46%	48%
% failures for which effective alternatives found (new PIs added)	49%	52%

Conclusions

- Models trained with large, representative datasets can predict virological response to cART accurately without a genotype.
- The results highlight viral load as the most important variable in modelling response
- Models are able to identify potentially effective 3-drug regimens comprising older drugs in a substantial proportion of failures
- This approach has potential for optimising antiretroviral therapy in resource-poor countries

Thanks to our data contributors

- BC Centre for Excellence in HIV/AIDS: *Richard Harrigan & Julio Montaner*
- NIAID: *Cliff Lane, Julie Metcalf, Robin Dewar*
- Gilead Sciences: *Michael Miller and Jim Rooney*
- The Italian HIV Cohort (University of Siena, Italy): *Maurizio Zazzi*
- US Military HIV Research Program: *Scott Wegner & Brian Agan*
- Hospital Clinic Barcelona: *Jose Gatell & Elisa Lazzari*
- Fundacion IrsiCaixa, Badelona: *Bonaventura Clotet & Lidia Ruiz*
- ICONA: *Antonella Monforte & Alessandro Cozzi-Lepri*
- Northwestern University, Chicago: *Rob Murphy & Patrick Milne*
- NCHECR, Sydney, Australia: *Sean Emery*
- Ramon y Cajal Hospital, Madrid, Spain: *María Jésus Pérez-Elías*
- Italian MASTER Cohort (c/o University of Brescia, Italy): *Carlo Torti*
- CPCRA: *John Bartlett, Mike Kozal, Jody Lawrence*
- Hôpital Timone, Marseilles, France*: Catherine Tamalet*
- ATHENA National Dutch database, Amsterdam: *Frank DeWolf & Joep Lange*
- Chelsea and Westminster Hospital, London: *Brian Gazzard, Anton Pozniak & Mark Nelson*
- Royal Free Hospital, London: *Anna Maria Geretti*
- Hospital of the JW Goethe University, Frankfurt: *Schlomo Staszewski*
- National Institute of Infectious Diseases, Tokyo: *Wataru Sugiura*
- Tibotec Pharmaceuticals: *Gaston Picchio and Marie-Pierre deBethune*

…and a special thanks to all their patients.

The RDI …

Brendan Larder

Dechao Wang

Daniel Coe