

**2018 update to the HIV-TRePS system: The development of new computational models to predict HIV treatment outcomes, with or without a genotype with enhanced usability for low-income settings**

Journal:	<i>Journal of Antimicrobial Chemotherapy</i>
Manuscript ID	JAC-2017-2100.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Revell, Andrew; RDI Wang, Dechao; RDI Perez-Elías, María; Ramon y Cajal Hospital, Infectious Diseases, IRYCIS Wood, Robin; University of Cape Town, Desmond Tutu HIV Centre Cogill, Dolphina; University of Cape Town, Desmond Tutu HIV Centre Tempelman, Hugo; Ndlovu Care Group, Hamers, Raph; University of Amsterdam, 5. Depts of Internal Medicine and Global Health, Academic Medical Centre Reiss, Peter; University of Amsterdam, Depts of Internal Medicine and Global Health; Stichting HIV Monitoring van Sighem, AI; Stichting HIV Monitoring Rehm, Catherine; National Institute of Allergy and Infectious Diseases (NIAID) Pozniak, A; Chelsea and Westminster Hospital, HIV; Montaner, Julio; 12. BC Centre for Excellence in HIV/AIDS, Lane, H; National Institute of Allergy and Infectious Diseases, Larder, Brendan; RDI
Keywords:	Antiretroviral therapy, Drug resistance, Genotype, Genotyping, HIV, HIV/AIDS, Medical decision making, Model, Modelling, Resistance

1     **2018 update to the HIV-TRePS system: The development of new computational**  
2     **models to predict HIV treatment outcomes, with or without a genotype with**  
3     **enhanced usability for low-income settings**

4  
5     **Short title:**

6     **2018 HIV-TRePS update: new models to predict HIV therapy response**  
7

8     Andrew D REVELL\*<sup>1</sup>, Dechao WANG<sup>1</sup>, Maria-Jesus PEREZ-ELIAS<sup>2</sup> Robin WOOD<sup>3</sup>,  
9     Dolphina COGILL<sup>3</sup>, Hugo TEMPELMAN<sup>4</sup>, Raph L HAMERS<sup>5</sup>, Peter REISS<sup>5,6</sup>, Ard I van  
10    SIGHEM<sup>6</sup>, Catherine A REHM<sup>7</sup>, Anton POZNIAK<sup>8</sup>, Julio SG MONTANER<sup>9</sup>, H Clifford  
11    LANE<sup>7</sup> & Brendan A LARDER<sup>1</sup> on behalf of the RDI study group†.

- 12           1. The HIV Resistance Response Database Initiative (RDI), London UK  
13           2. Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain  
14           3. Desmond Tutu HIV Centre, University of Cape Town, South Africa  
15           4. Ndlovu Care Group, Elandsdoorn, South Africa  
16           5. Depts of Internal Medicine and Global Health, Academic Medical Centre  
17           of the University of Amsterdam, Amsterdam Institute for Global Health  
18           and Development, Amsterdam, The Netherlands  
19           6. Stichting HIV Monitoring, Amsterdam, The Netherlands  
20           7. National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA  
21           8. Chelsea and Westminster Hospital, London, UK  
22           9. BC Centre for Excellence in HIV/AIDS, Vancouver, Canada

23    Correspondence to Andrew D. Revell, 14 Union Square, London N1 7DH, UK.

24 Tel: +44 207 226 7314; fax: +44 207 226 7314; e-mail: andrewrevell@hivrdi.org

25

26 **Footnote:**

27 †Members are listed in the Acknowledgements section.

28

## 29    **Abstract**

### 30    **Objective**

31    Optimising antiretroviral drug combination on an individual basis can be challenging  
32    particularly in settings with limited access to drugs and genotypic resistance testing.  
33    Here we describe our latest computational models to predict treatment responses,  
34    with or without a genotype, and compare their predictive accuracy with that of  
35    genotyping.

36

### 37    **Methods**

38    Random forest models were trained to predict the probability of virological response  
39    to a new therapy introduced following virological failure using up to 50,000  
40    treatment change episodes (TCEs) without a genotype, and 18,000 TCEs *including*  
41    genotypes. Independent data sets were used to evaluate the models. This study  
42    tested the effects on model accuracy of relaxing the baseline data timing windows,  
43    the use of a new filter to exclude probable non-adherent cases and the addition of  
44    maraviroc, tipranavir and elvitegravir to the system.

### 45    **Results**

46    The no-genotype models achieved area under the receiver-operator characteristic  
47    curve (AUC) values of 0.82 and 0.81 using the standard and relaxed baseline data  
48    windows respectively. The genotype models achieved AUC values of 0.86 with the  
49    new non-adherence filter and 0.84 without.

50    Both sets of models were significantly more accurate than genotyping with rules-  
51    based interpretation, which achieved AUC values of only 0.55-0.63, and marginally

52 more accurate than previous models. The models were able to identify alternative  
53 regimens that were predicted to be effective for the vast majority of cases where the  
54 new regimen prescribed in the clinic failed.

## 55 **Conclusions**

56 These latest global models predict treatment responses accurately even without a  
57 genotype and have the potential to help optimise therapy, particularly in resource-  
58 limited settings.

## 59    **Introduction**

60    The development of approximately 30 HIV drugs acting at six different points in the  
61    virus lifecycle and the expansion of access to therapy around the world is great  
62    success story<sup>1</sup>. The current UNAIDS target for 2020 is for 90% of infected people to  
63    be diagnosed, 90% of them to be on therapy with 90% of those treated having  
64    suppressed virus ("90-90-90"). The last target is critical in order not only to prevent  
65    disease progression, morbidity and mortality but to decrease the spread of the  
66    virus.<sup>2,3</sup> A major threat to this is the development of HIV drug resistance, often  
67    linked to poor adherence and interruptions to drug supplies in some settings.

68    A recent report from the WHO and others showed that the prevalence of HIV drug  
69    resistance among patients in public health antiretroviral treatment programmes has  
70    increased from 11% to 29% since the global expansion of ART to low and middle  
71    income countries (LMICs) began in 2001.<sup>1</sup>

72    When treatments fails, the combination of antiretroviral agents should be changed  
73    in order to re-suppress the virus. In most well-resourced countries the selection of a  
74    new combination is individualized by expert physicians using information including  
75    the patient's treatment history and the results of a genotypic resistance test.<sup>4-6</sup>  
76    However, resistance testing is relatively expensive and only moderately predictive of  
77    response to treatment.<sup>7</sup>

78    The challenge of individually optimized drug selection in LMICs is even greater as  
79    resistance tests are typically unavailable or unaffordable and drug options are  
80    limited.<sup>8</sup> In the absence of routine viral load monitoring, therapy failure is often  
81    detected late and regimen switch decisions based on standard protocols rather than

82 individualized. The result can be sub-optimal regimen selection, failure to achieve  
83 viral re-suppression and further resistance selection, which may limit future  
84 therapeutic options and can be transmitted to others.<sup>9</sup>

85 The HIV Resistance Response Database Initiative (RDI) has collected biological,  
86 clinical and treatment outcome data for more than 200,000 HIV-1 patients around  
87 the world over a period of 16 years. From these data, we have used machine  
88 learning to develop models to predict HIV-1 treatment outcomes and to identify  
89 optimal, individualized therapies.<sup>10-15</sup> We have developed models that use  
90 information from genotypic resistance tests in their predictions and others that do  
91 not. The most recent models, developed using large datasets from around the world  
92 then tested with independent test sets predicted virological response with an overall  
93 accuracy of around 80% with a genotype and 74% without.<sup>14,15</sup>

94 The models are used to power an online treatment decision support tool, the HIV  
95 Treatment Response Prediction System (HIV-TRePS). To keep this system as current  
96 as possible in terms of the inclusion of new drugs and reflection of current clinical  
97 practice it is essential that new models are regularly developed using the latest data.

98 Here we report the development of two new sets of random forest (RF) models that  
99 estimate the probability of combinations of antiretroviral drugs reducing the plasma  
100 viral load to undetectable (<50 copies HIV RNA/mL):

101 1. Models that do not require a genotype for their predictions (no-genotype, or NG  
102 models), trained using a large global data set and intended for use in LMICs  
103 without access to genotyping. We compared standard models (NG1) with

104 experimental models developed using new highly permissive data inclusion  
105 criteria (NG2) to increase utility in LMICs where clinic visits can be infrequent.

106 2. Global models that use a viral genotype in their predictions (global genotype, or  
107 G models). These were developed using data screened for likely non-adherence  
108 using an experimental filter. Cases of discordance between virological failure  
109 observed in the clinic and predictions of response by both our current models  
110 and genotyping with rules based interpretation were excluded (G2) and the  
111 resultant models compared to 'standard' models (G1).

112 3. For the first time there were sufficient data without genotypes for the NG  
113 models to be trained to predict outcomes for three drugs not previously covered,  
114 tipranavir, maraviroc and elvitegravir.

115 The accuracy of all the models was ascertained and evaluated as potential tools to  
116 support optimized, individualized treatment decision-making in the RDI's HIV-TRePS  
117 system. This paper represents the latest update alluded to in our previous  
118 publications of modeling.<sup>13-15</sup>

119

120

## 121 **Methods**

### 122 ***Clinical data***

123 Treatment change episodes (TCEs) were collected from cases where antiretroviral  
124 therapy was changed following virological failure.<sup>10</sup> TCEs for development of the NG  
125 models had all the following data available: On-treatment baseline plasma viral load  
126 (obtained  $\leq 8$  weeks prior to treatment change for the standard models, NG1 and  $\leq 12$



127 weeks for the experimental models, NG2); on treatment baseline CD4 cell count ( $\leq 12$   
128 weeks prior to treatment change for NG1 and  $\leq 16$  weeks for the NG2); the drugs in  
129 use prior to the change; antiretroviral treatment history; drugs in the new regimen;  
130 follow-up plasma viral load obtained 4-52 weeks following introduction of the new  
131 regimen and time to that follow-up. A similar extraction was subsequently  
132 performed for TCEs that also had an on treatment genotype (protease and RT  
133 sequence  $\leq 12$  weeks prior to treatment change) for the development of the  
134 genotype models.

135 The TCEs were censored using rules established in previous studies and published in  
136 detail elsewhere.<sup>16</sup>

#### 137 ***Data partition for models without genotypes (NG)***

138 The qualifying TCEs were partitioned using methods described elsewhere.<sup>11,16</sup> The  
139 partition scheme is illustrated in Figure 1. For the NG2 models, with the expanded  
140 baseline data windows, a training set of 50,270 TCEs and an independent test set of  
141 3,000 TCEs, one each from 3,000 patients not represented in the training set were  
142 obtained. For NG1 models, the standard baseline data windows were applied to the  
143 NG2 training and test sets resulting in training and test sets reduced to 43,239 and  
144 2,500 TCEs.

145

#### 146 ***Data partition for models with genotypes (G)***

147 The data that included baseline genotypes were partitioned into a master pool of  
148 18,242 training TCEs and 1,000 test TCEs. The data were screened for possible non-

149 adherence using our standard filter (which excludes cases with a baseline viral load  
150 of  $\leq 3.0 \log_{10}$  HIV RNA and an increase in viral load of  $\geq 2.0$  following the introduction  
151 of a new regimen selected with a recent genotype available). This resulted in a  
152 training set of 18,188 and a test set of 997. The master pool of TCEs was then  
153 screened using an experimental filter for possible non-adherence. Cases where the  
154 new regimen used in the clinic failed to achieve virological response were passed  
155 through HIV-TRePS. Those predicted to respond (follow-up viral load  $\leq 50$  copies)  
156 were extracted and genotype sensitivity scores obtained using the Stanford HIVdb,  
157 REGA and ANRS interpretation systems. Cases with a GSS of  $\geq 2$  (two or more active  
158 drugs) in all three systems were then excluded. This removed approximately 5% of  
159 the TCEs, resulting in a training set of 17,378 and test set of 940.

160

### 161 **Computational model development**

162 The two NG training sets of TCEs were used to train two committees of 5 RF models,  
163 to estimate the probability of the follow-up viral load being less than 50 copies/mL,  
164 using methodology described elsewhere.<sup>11,13</sup> The following 47 input variables were  
165 used (new variables underlined):

166 Baseline viral load ( $\log_{10}$  copies HIV RNA/mL); baseline CD4 count ( $\text{cells}/\text{mm}^3$ );  
167 treatment history - 22 binary variables coding for experience of zidovudine,  
168 didanosine, stavudine, abacavir, lamivudine, emtracitabine, tenofovir DF, efavirenz,  
169 nevirapine, etravirine, indinavir, nelfinavir, saquinavir, amprenavir, fos-amprenavir,  
170 lopinavir, atazanavir, darunavir, enfuvirtide, raltegravir, tipranavir, maraviroc;  
171 antiretroviral drugs in the new regimen - 23 variables as above with the addition of

172 elvitegravir; time to follow-up (days). The output variable was the follow-up viral  
173 load as a binary variable:  $\leq 1.7$  log or 50 copies/mL = 1 (response) and  $>1.7$  log or 50  
174 copies/mL = 0 (failure).

175  
176 The genotype models used 105 input variables including the above but without  
177 raltegravir as a historical drug and without maraviroc or tipranavir in the new  
178 regimen because of insufficient data with genotypes. Historical maraviroc was a new  
179 variable. In addition, the following 62 mutations, detected in the baseline genotype  
180 were used: HIV reverse transcriptase mutations (n=33): M41L, E44D, A62V, K65R,  
181 D67N, 69 insert, T69D/N, K70R, L74V, V75I, F77L, V90I, A98G, L100I, L101I/E/P,  
182 K103N, V106A/M, V106I, V108I, Y115F, F116Y, V118I, I38A/G/K, Q151M, V179D/F/T,  
183 Y181C/I/V, M184V/I, Y188C/L/H, G190S/A, L210W, T215F/Y, K219Q/E, P236L;  
184 protease mutations (n=29): L10F/I/R/V, V11I, K20M/R, L24I, D30N, V32I, L33F, M36I,  
185 M36L/V, M46I/L, I47V, G48V, I50V, I50L, F53L, I54 (any change), 58E, L63P, A71(any  
186 change), G73(any change), T74P, L76V, V77I, V82A/F/S, V82T, I84V/A/C, N88D/S,  
187 L89V, L90M). The mutations were selected on the basis of the IAS-USA mutation list  
188 as well as previous modelling studies.<sup>17</sup>

189

## 190 **Validation and independent testing**

191 Each of the four committees of 5 RF models was developed using a 5 x cross  
192 validation scheme.<sup>11,16</sup> For each partition the model's estimates of the probability of  
193 response for the validation cases was compared to the actual response observed in  
194 the clinic and the best performing model selected for the final committee. For each of

195 the five final models, the optimum operating point (OOP) was identified (the cut-off  
196 for the probability of response being classed as response versus failure that gave the  
197 best performance overall).

198 The performance of the models as predictors of response was then evaluated using the  
199 independent test cases. The models' estimates of the probability of response and the  
200 responses observed in the clinics for these cases were used to plot receiver-operator  
201 characteristic (ROC) curves and assess the area under the ROC curve (AUC). In addition, the  
202 average OOP, derived during cross validation, was used to obtain the overall accuracy, the  
203 sensitivity and the specificity of the models.

#### 204 **Comparison of the accuracy of the models versus rules-based interpretation of the** 205 **genotype**

206 Genotypic sensitivity scores (GSS) were obtained for test cases with genotypes that  
207 could be fully interpreted by three rules-based genotype interpretation systems in  
208 common use: ANRS, REGA and Stanford HIVdb). The three systems were accessed  
209 online on 07/09/2017 and the GSS calculated by adding the score for each drug in  
210 the regimen, with full susceptibility scored as 1, partial as 0.5 and no response as 0.  
211 These scores were then used as predictors of response and the performance  
212 compared to that of the models.<sup>16</sup>

213

#### 214 **In silico analysis to evaluate the potential of the models to help avoid treatment** 215 **failure**

216 In order to evaluate further the potential clinical utility of the models, we assessed their  
217 ability to identify alternative, practical regimens that were predicted to be effective  
218 (probability of virological response above the OOP), or more likely to be effective than the  
219 regimens introduced in the clinic. Lists of regimens in regular clinical use were identified  
220 from the RDI database. The baseline data for all test TCEs where the new regimen  
221 comprised three or more drugs were entered into the models and predictions obtained for  
222 the regimens on the drug lists that had no more drugs than the regimen used in the clinic.  
223 Since the NG models are used primarily in LMICs, we wanted to avoid modelling regimens  
224 that are unavailable in such settings, as this could over-estimate the system's utility. The  
225 analysis was therefore repeated using test cases from sub-Saharan countries only and  
226 modelling alternative regimens comprising only those drugs that were in use in those  
227 countries at the time of data collection.

## 228 Results

### 229 Characteristics of the datasets

230 The baseline, treatment and response characteristics of the data sets are  
231 summarised in Tables 1 and 2. The training sets for NG models comprised 43,239  
232 TCEs using the standard baseline data windows and 50,270 using the expanded  
233 baseline data windows. The data sets have very comparable baseline data with a  
234 median plasma viral load of approximately 3.8 log<sub>10</sub> and CD4 count of approximately  
235 260. The median number of previous drugs used in the patients' treatment was 4-5,  
236 with almost all exposed to NRTIs, around two-thirds having been exposed to NNRTIs  
237 and two-thirds to PIs. There was a broad range of new regimens represented in the

238 data, the most common being two NRTIs and one PI (32-36%), followed by two NRTIs  
239 and an NNRTI (18-23%).

240

241 The characteristics of the TCEs with genotypes are summarised in Table 2. The  
242 training sets comprised 18,188 TCEs using the standard non-adherence filter and  
243 17,378 using the new filter. The sets have very comparable baseline data with a  
244 median plasma viral load of approximately 4.3 log<sub>10</sub> (about half a log higher than the  
245 NG data) and median CD4 count of approximately 230, slightly lower than the NG  
246 data. The treatment history was similar to that of the NG data as was the range of  
247 new regimens other than somewhat fewer patients changing to NNRTI-based  
248 regimens (approximately 10% versus 20%).

249

#### 250 **Results of the modelling without a genotype**

251 The performance characteristics from the ROC curves of the models during cross-  
252 validation and independent testing are summarized in Table 3. The NG1 models  
253 achieved AUC values during cross-validation of 0.83 to 0.84, with a mean of 0.84. The  
254 overall accuracy was 77% to 78% (mean=77%), the sensitivity was 71% for all five  
255 models and the specificity ranged from 80% to 81% (mean=80%). The OOP was 0.42-  
256 0.43.

257 The NG2 models achieved very similar results with AUC values during cross  
258 validation of 0.83 to 0.84 and a mean of 0.83. Overall accuracy ranged from 76% to  
259 77% (mean = 76%), sensitivity again was 71% and specificity ranged from 79% to 80%  
260 (mean = 80%).

261

**262 Independent testing**

263 The NG1 models achieved an AUC of 0.82 in independent testing (Figure 1). Overall  
264 accuracy was 75%, sensitivity 72% and specificity 77%. The NG2 models achieved an  
265 AUC value of 0.81, overall accuracy of 75%, sensitivity was 73% and specificity 76%.  
266 The performance of the two sets of models during independent testing was not  
267 significantly different ( $p=0.84$ ).

268 When Models NG2 were tested using only the 500 test cases with baseline data that  
269 fell outside of the standard windows, the AUC was 0.79, overall accuracy 73%,  
270 sensitivity and specificity each 73%. There were no significant differences between  
271 the performance of NG2 models with those cases with baseline data inside versus  
272 outside the standard windows ( $p=0.29$ ).

273 When the two sets of models were tested only with those cases involving each of the  
274 three new drugs (50 cases in each subset) the AUC values ranged from 0.75 to 0.89.

275

**276 Comparing the predictive accuracy of the models versus genotyping**

277 Of the 3,000 TCEs in the global no-genotype test set, 634 had genotypes available  
278 suitable for full interpretation by the three interpretation systems. The ROC curves  
279 are presented in Figure 2, alongside those for the models. The AUC values for the  
280 GSS were 0.56 (ANRS), 0.58 (Stanford HIVdb) and 0.55 (Rega) (Table 5). All were  
281 significantly less accurate predictors of virological response than the NG models,  
282 both sets of which achieved an AUC of 0.81 for these cases ( $p<0.0001$ ).

## 283     **Results of the modelling with a genotype**

284     The performance characteristics from the ROC curves of these models during cross-  
285     validation and independent testing are summarized in Table 4. The 5 G1 models  
286     achieved AUC values during cross-validation ranging from 0.84 to 0.86, with a mean  
287     of 0.86. Overall accuracy was 78% to 79% (mean=79%), sensitivity 71% to 73%  
288     (mean=73%) and specificity 78% to 82% (mean=81%). The OOP was 0.42.

289     The 5 G2 models (using data with the new non-adherence filter) achieved AUC  
290     values during cross-validation ranging from 0.86 to 0.89, with a mean of 0.88. Overall  
291     accuracy was 79% to 82% (mean=80%), sensitivity 77% to 81% (mean=79%) and  
292     specificity 79% to 82% (mean=81%). The OOP was again 0.42.

293

## 294     **Independent testing**

295     When tested with the independent G1 test cases using the OOP developed in cross  
296     validation, the G1 models achieved an AUC of 0.84 (Figure 3). The overall accuracy  
297     was 76%, sensitivity 72% and specificity 80%. Models G2 achieved an AUC value  
298     during testing with the G2 test set of 0.86, overall accuracy of 79%, sensitivity of 74%  
299     and specificity of 83%. Again, the G2 models performance was slightly better than  
300     G1 but there was no statistically significant difference between the two sets of  
301     models ( $p=0.25$ ).

302     When models G1 were tested using the G2 test set, performance improved slightly  
303     (AUC from 0.84 to 0.85 and overall accuracy from 76% to 77%) but was not as good  
304     as the G2 models. Conversely the performance of the G2 models worsened when



305 tested with the G1 test set (AUC reduced from 0.86 to 0.83 and overall accuracy  
306 from 79% to 76%), but remained comparable to the performance of the G1 models.

### 307 **Comparing the predictive accuracy of the G2 genotype models versus genotyping**

308 Genotypic sensitivity scores were generated for 856 test TCEs where the drugs in the  
309 new regimens were fully covered by the interpretation systems. The genotype  
310 systems achieved AUC values of 0.60-0.63, compared with 0.86 using the G2 models  
311 and 0.84 for G1 (Figure 3). The overall accuracy figures were 57-59% for genotyping  
312 compared with 80% for the G2 models and 77% for G1. All three genotype  
313 interpretation systems were significantly worse at predicting responses than both  
314 sets of models ( $p < 0.00001$ ).

### 315 **In silico analysis**

316 The NG models were able to identify alternative regimens that were predicted to be  
317 effective for 97% (NG1) to 98% (NG2) of cases (Table 6). They were able to identify  
318 alternative regimens comprising only those drugs in common use with a higher  
319 probability of response (but not necessarily above the response classification  
320 threshold of the models) for all cases. With the cases where the new regimen failed  
321 in the clinic, the models were able to identify alternatives that were predicted to be  
322 effective in 96% and with a higher probability of response in 100%.

323 Using only locally available drugs (3TC, ABC, AZT, ddI, EFV, FTC, LPV, NVP, RAL, TDF) for 450  
324 cases from sub-Saharan Africa, the NG models were able to identify alternative  
325 regimens that were predicted to be effective for 95% of cases and 92-93% for cases  
326 where the new regimen failed in the clinic.

327 The genotype models identified alternative regimens that were predicted to give a  
328 response for 93% of the test cases and regimens with a higher probability of  
329 response for 99.9% (Table 6). For patients who experienced virological failure in the  
330 clinic, the models identified alternatives that were predicted to give a response for  
331 90% and with a higher probability of response than the regimen in the clinic for all  
332 100%.

333

### 334 Discussion

335 These latest computational models, developed using our largest databases are the  
336 most accurate predictors of response to combination antiretroviral therapy to date.  
337 They include, for the first time, tipranavir, maraviroc and elvitegravir.

338 Both sets of models achieved AUC values over 0.80 in cross validation and  
339 independent testing and there was no significant difference in accuracy between  
340 those that use a genotype and those that do not. The results replicated and  
341 reinforced previous findings that our models are substantially more accurate  
342 predictors of virological response to combination therapy than viral genotyping with  
343 rules-based interpretation.<sup>14,15</sup> It is encouraging that this superiority was maintained  
344 despite eliminating a number of cases where the GSS and the response observed in  
345 the clinic were discordant, in an attempt exclude non-adherent patients.

346 The expanded 'windows' for baseline data make these models more practical for use  
347 in LMICs where visits for laboratory monitoring are relatively infrequent. Moreover,  
348 the broad range of settings represented in the study data suggests that these  
349 findings and the potential utility of these models are highly generalizable.

350 Nevertheless, in some LMICs viral loads or CD4 counts may lay outside even these  
351 extended windows, if available at all, preventing use of the system. Given the size of  
352 the RDI database, research looking at models that can accept missing values is  
353 warranted.

354 The use of a new more stringent filter for presumed non-adherence removed a  
355 greater proportion of available TCEs than the standard filter and led to a small  
356 numerical increase in performance for the genotype models.

357 The NG models presented here can predict outcomes to 23 different drugs, including  
358 some relatively recently approved inhibitors that are not routinely available in  
359 LMICs. Users of the HIV-TRePS system are able to exclude any drugs that are not  
360 locally available from the modelling. The *in silico* results for cases from LMICs, using a  
361 highly restricted list of locally available drugs, demonstrated the potential of the  
362 models to improve virological response rates nevertheless, underlining their  
363 applicability for LMICs.

364 A key input variable for these models was the plasma viral load, which studies have  
365 shown to be important to the accuracy of the models.<sup>18</sup> Although viral load  
366 monitoring is not routine in LMICs, it is now recommended in WHO guidelines for  
367 monitoring antiretroviral therapy response.<sup>19</sup> Accurate statistics on viral load  
368 monitoring in LMICs are scarce. However, a recent study of its scale-up in sub-  
369 Saharan Africa showed the percentage of patients with viral loads ranged from 3%  
370 (Cote-d'Ivoire and Tanzania) to 96% in Namibia. Of 11 million patients on  
371 antiretroviral therapy in the region, 5 million were estimated to have access to viral

372 load monitoring.<sup>20</sup> As technological advances enable lower costs and point of care  
373 testing, the use of viral load is likely to increase.<sup>21,22</sup>

374 The study has some limitations. Firstly, it was retrospective and no firm claims can be  
375 made for the clinical benefit of using the system as a treatment support tool. This  
376 would require large, prospective clinical trials, for example comparing outcomes for  
377 patients with a change to their treatment managed under standard of care (SoC)  
378 versus SoC plus the HIV-TRePS report.

379 The models described here were trained to estimate the probability of response to  
380 therapy using a definition of response of <50 copies HIV RNA/mL. While recent data  
381 strongly suggest that low-level viraemia predicts virological failure, differences  
382 persist in the definition of virological response used in the clinic.<sup>23</sup> The US AIDS  
383 Clinical Trials Group (ACTG) define virological failure as a confirmed viral load >200  
384 copies/mL.<sup>6</sup> The World Health Organisation, however, defines virological failure as  
385 persistent plasma RNA levels ≥1000 copies/mL after 3 months with adherence  
386 support.<sup>19</sup> Other groups are using 400 copies/mL.<sup>24</sup>

387 The models may predict that a certain combination of drugs is likely to fail (viral load  
388 >50 copies) with no indication of the probability of the viral load being below a  
389 different cut-off e.g. 1,000/mL copies. Studies are now ongoing to develop models  
390 that predict absolute viral load value over time following a treatment change.

391

392 **Conclusions**

393 Computational models developed using large, heterogeneous data sets with  
394 relatively permissive rules governing the timing of baseline data are highly accurate

395 predictors of virological response to combination antiretroviral therapy, even  
396 without a genotype. Such models are of enhanced utility in settings with infrequent  
397 laboratory monitoring.

398 Attempts to remove any possible contamination of training with data from non-  
399 adherent patients continue. A new filter for presumed non-adherence removed  
400 around 5% of training data and led to a very small increase in accuracy.

401 The models were able to predict responses to tipranavir, maraviroc and elvitegravir  
402 for the first time and with accuracy comparable to that of other antiretrovirals, again  
403 expanding the utility of the system.

404 These latest models, are better predictors of response to therapy than genotyping  
405 with rules-based interpretation, even when those models do not use a genotype for  
406 their predictions. Since use of these models is free of charge, this suggests that  
407 scarce funds in LMICs would be better spent on antiretroviral drugs and viral load  
408 testing than on genotyping. This would enable a greater range of treatments to be  
409 offered, failure to be detected early and optimal, individualised treatment change  
410 decisions made using the models.

411 Full validation of this approach as a clinical tool would require a prospective,  
412 controlled clinical trial. Nevertheless, the results suggest that these models have the  
413 potential to reduce virological failure and improve patient outcomes in all parts of  
414 the world, with particular utility in LMICs. The use by clinicians of this tool to support  
415 optimised treatment decision-making in the absence of resistance tests could also  
416 combat the development of drug resistance and its contribution to treatment failure,  
417 disease progression and onward viral transmission.

418

419 The global models described in this paper are freely available to use online through  
420 the HIV-TRePS system at <http://www.hivrddi.org/treps>.

421

## 422 **Acknowledgements**

423 **RDI data and study group:** The RDI wishes to thank all the following individuals and  
424 institutions for providing the data used in training and testing its models:

425 **Cohorts:** Peter Reiss and Ard van Sighem (ATHENA, the Netherlands); Julio Montaner  
426 and Richard Harrigan (BC Center for Excellence in HIV & AIDS, Canada); Tobias Rinke  
427 de Wit, Raph Hamers and Kim Sigaloff (PASER-M cohort, The Netherlands); Brian  
428 Agan, Vincent Marconi and Scott Wegner (US Department of Defense); Wataru  
429 Sugiura (National Institute of Health, Japan); Maurizio Zazzi (MASTER, Italy); Rolf  
430 Kaiser and Eugen Schuelter (Arevir Cohort, Köln, Germany); Adrian *Streinu-Cercel*  
431 (National Institute of Infectious Diseases Prof.Dr. Matei Balş, Bucharest, Romania);  
432 Gerardo Alvarez-Uria (VFHCS, India), Maria-Jesus Perez-Elias, (CORIS, Spain), Tulio de  
433 Oliveira, (SATuRN, South Africa).

434

435 **Clinics:** Jose Gatell and Elisa Lazzari (University Hospital, Barcelona, Spain); Brian  
436 Gazzard, Mark Nelson, Anton Pozniak and Sundhiya Mandalia (Chelsea and  
437 Westminster Hospital, London, UK); Colette Smith (Royal Free Hospital, London, UK);  
438 Lidia Ruiz and Bonaventura Clotet (Fundacion Irsi Caixa, Badelona, Spain); Schlomo  
439 Staszewski (Hospital of the Johann Wolfgang Goethe-University, Frankfurt,  
440 Germany); Carlo Torti (University of Brescia); Cliff Lane, Julie Metcalf and Catherine A  
441 Rehm (National Institutes of Health Clinic, Rockville, USA); Maria-Jesus Perez-Elias  
442 (Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain); Stefano Vella and  
443 Gabrielle Dettorre (Sapienza Univiersity, Rome, Italy); Andrew Carr, Richard Norris  
444 and Karl Hesse (Immunology B Ambulatory Care Service, St. Vincent's Hospital,  
445 Sydney, NSW, Australia); Dr Emanuel Vlahakis (Taylor's Square Private Clinic,  
446 Darlinghurst, NSW, Australia); Hugo Tempelman and Roos Barth (Ndlovu Care Group,

447 Elandsdoorn, South Africa), Robin Wood, Carl Morrow and Dolphina Cogill (Desmond  
448 Tutu HIV Centre, University of Cape Town, South Africa); Chris Hoffmann (Aurum  
449 Institute, Johannesburg, South Africa and Johns Hopkins University, Boston, USA;  
450 Luminita Ene (“Dr. Victor Babes” Hospital for Infectious and Tropical Diseases,  
451 Bucharest, Romania); Gordana Dragovic (University of Belgrade, Belgrade, Serbia);  
452 Ricardo Diaz and Cecilia Sucupira (Federal University of Sao Paulo, Sao Paulo Brazil);  
453 Omar Sued and Carina Cesar (Fundación Huésped, Buenos Aires, Argentina); Juan  
454 Sierra Madero (Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran,  
455 Mexico City, Mexico); Pachamuthu Balakrishnan and Shanmugam Saravanan (YRG  
456 Care, Chennai, India).

457

458 **Clinical trials:** Sean Emery and David Cooper (CREST); Carlo Torti (GenPherex); John  
459 Baxter (GART, MDR); Laura Monno and Carlo Torti (PhenGen); Jose Gatell and  
460 Bonventura Clotet (HAVANA); Gaston Picchio and Marie-Pierre deBethune (DUET 1 &  
461 2 and POWER 3); Maria-Jesus Perez-Elias (RealVirfen); Sean Emery, Paul Khabo and  
462 Lotty Ledwaba (PHIDISA).

463

#### 464 **Funding**

465 This project has been funded in whole or in part with federal funds from the National  
466 Cancer Institute, National Institutes of Health, under Contract No.  
467 HHSN261200800001E. This research was supported by the National Institute of  
468 Allergy and Infectious Diseases. The content of this publication does not necessarily  
469 reflect the views or policies of the Department of Health and Human Services, nor  
470 does mention of trade names, commercial products, or organizations imply  
471 endorsement by the U.S. Government.

472

473     **Transparency statement**

474     None to declare

475

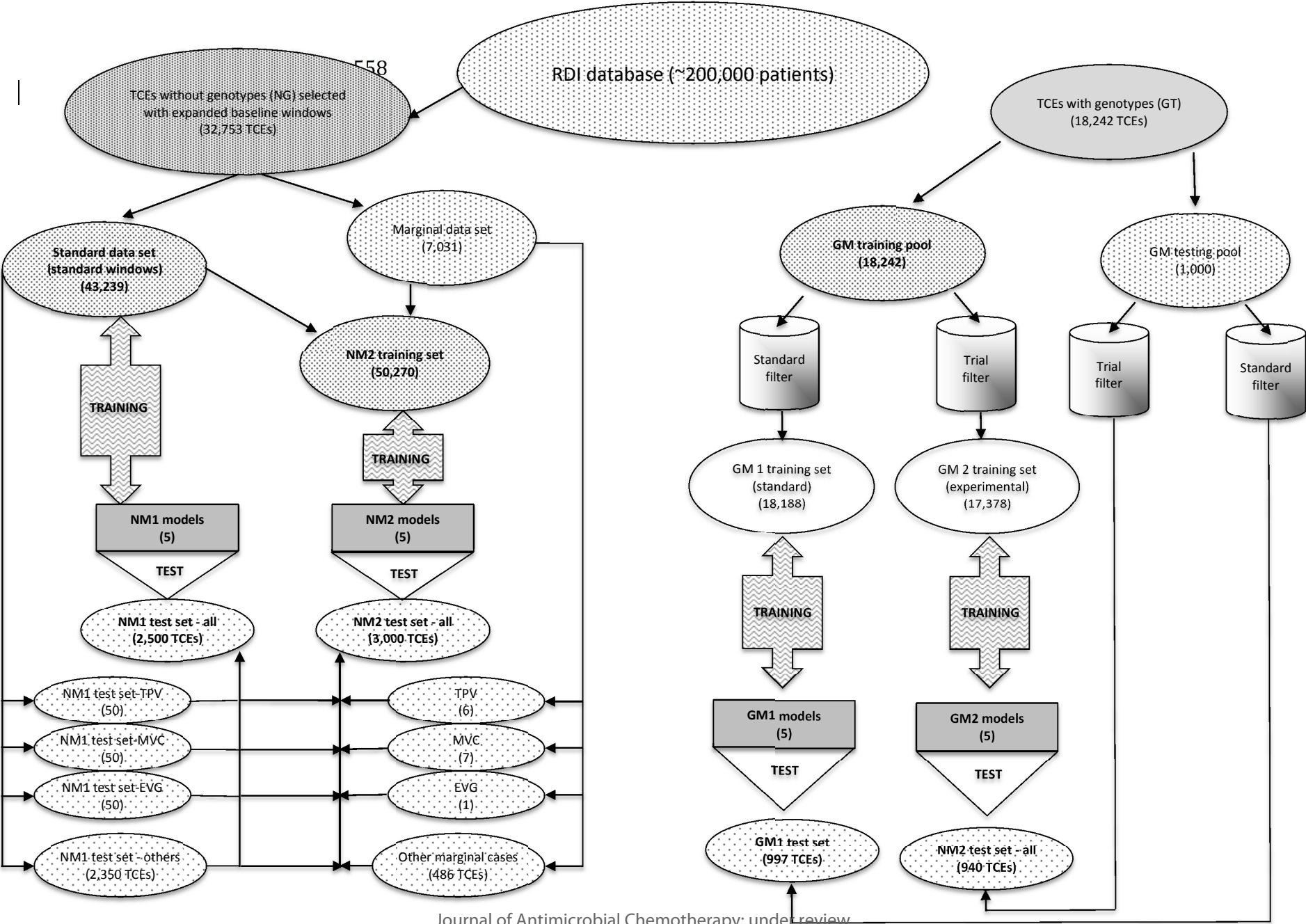


476 **References**

- 477 1. World Health Organization. HIV drug resistance report 2017 ([www.who.int/hiv/pub/drug-resistance/hivdr-report-2017/en/](http://www.who.int/hiv/pub/drug-resistance/hivdr-report-2017/en/)).
- 478
- 479 2. Montaner JS, Lima VD, Harrigan PR, *et al.* Expansion of HAART Coverage Is
- 480 Associated with Sustained Decreases in HIV/AIDS Morbidity, Mortality and
- 481 HIV Transmission: The “HIV Treatment as Prevention” Experience in a
- 482 Canadian Setting. *PLoS ONE* 2014; 9: e87872.
- 483 <https://doi.org/10.1371/journal.pone.0087872>
- 484 3. Nkambule R, Nuwagaba-Biribonwoha H, Mnisi Z, *et al.* Substantial progress in
- 485 confronting the HIV epidemic in Swaziland: first evidence of national impact.
- 486 In: *Abstracts of the 9th International Conference on IAS Science, Paris, 2017*.
- 487 Abstract MOAX0204LB, p 564. International AIDS Society, Geneva,
- 488 Switzerland.
- 489 4. Günthard HF, Aberg JA, Eron JJ *et al.* Antiretroviral Treatment of Adult HIV
- 490 Infection: 2014 Recommendations of the International Antiviral Society–USA
- 491 Panel. *JAMA* 2014; **312**: 410-425.
- 492 5. Writing Group, Williams I, Churchill D *et al.* British HIV Association guidelines
- 493 for the treatment of HIV-1-positive adults with antiretroviral therapy 2012
- 494 (Updated November 2013) *HIV Medicine* 2014; **15**:1–85.
- 495 6. *Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for*
- 496 *the use of antiretroviral agents in HIV-1-infected adults and adolescents*.
- 497 Department of Health and Human Services. Available at
- 498 <http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf>.
- 499 7. Degruittola V, Dix L, D'Aquila R *et al.* The relation between baseline HIV drug
- 500 resistance and response to antiretroviral therapy: re-analysis of retrospective
- 501 and prospective studies using a standardized data analysis plan. *Antivir Ther*
- 502 2000; **5**: 41-48.
- 503 8. *South African Drug Resistance Guidelines*. Available at:
- 504 [http://www.sahivsoc.org/upload/documents/HIVDR Strategy.pdf](http://www.sahivsoc.org/upload/documents/HIVDR%20Strategy.pdf)

- 505 9. Gupta RK, Hill A, Sawyer AW *et al.* Virological monitoring and resistance to  
506 first-line highly active antiretroviral therapy in adults infected with HIV-1  
507 treated under WHO guidelines: a systematic review and meta-analysis.  
508 *Lancet Infect Dis* 2009; **9**: 409-17.
- 509 10. Larder B, Wang D, Revell A *et al.* The development of artificial neural  
510 networks to predict virological response to combination HIV therapy. *Antivir.*  
511 *Ther* 2007; **12**: 15-24.
- 512 11. Wang D, Larder B, Revell A *et al.* A comparison of three computational  
513 modelling methods for the prediction of virological response to combination  
514 HIV therapy. *Artif Intell Med* 2009; **47**: 63-74.
- 515 12. Larder BA, Revell A, Mican JM *et al.* Clinical evaluation of the potential utility  
516 of computational modelling as an HIV treatment selection tool by physicians  
517 with considerable HIV experience. *AIDS Patient Care STDS* 2011; **25**: 29-36.
- 518 13. Revell AD, Wang D, Wood R *et al.* Computational models can predict  
519 response to HIV therapy without a genotype and may reduce treatment  
520 failure in different resource-limited settings. *J Antimicrob. Chemother* 2013;  
521 **68**:1406-1414.
- 522 14. Revell AD, Wang D, Wood R *et al.* An update to the HIV-TRePS system: the  
523 development of new computational models that do not require a genotype  
524 to predict HIV treatment outcomes. *J Antimicrob. Chemother* 2014; **69**:1104-  
525 1110.
- 526 15. Revell AD, Wang D, Wood R *et al.* An update to the HIV-TRePS system: the  
527 development and evaluation of new global and local computational models  
528 to predict HIV treatment outcomes, with or without a genotype. *J Antimicrob.*  
529 *Chemother* 2016; **71**:2928-2937.
- 530 16. Revell AD, Wang D, Boyd MA *et al.* The development of an expert system to  
531 predict virological response to HIV therapy as part of an online treatment  
532 support tool. *AIDS* 2011; **25**:1855-1863.
- 533 17. Wensing AM, Calvez V, Huldrych F *et al.* 2015 Update of the Drug Resistance  
534 Mutations in HIV-1. *Top Antivir Med* 2015; **23**:132-141.

- 535 18. Revell AD, Wang D, Harrigan R *et al.* Modelling response to HIV therapy  
536 without a genotype – an argument for viral load monitoring in resource-  
537 limited settings. *J Antimicrob Chemother* 2010; **65**:605-7.
- 538 19. WHO. Consolidated Guidelines on the use of Antiretroviral Drugs for Treating  
539 and Preventing HIV Infection. Recommendations For a Public Health  
540 Approach – Second Edition. Geneva: WHO, 2016.  
541 <http://www.who.int/hiv/pub/guidelines/arv2013/download/en/>
- 542 20. Lecher S, Ellenberger D, Kim AA *et al.* Scale up of HIV viral load monitoring –  
543 seven sub-Saharan African countries. *MMWR* 2015; **64**:1281-1304
- 544 21. Stevens WS, Scott LE, Crowe SM. Quantifying HIV for monitoring  
545 antiretroviral therapy in resource-poor settings. *J Infect Dis* 2010; **201**: S16–  
546 26
- 547 22. Roberts T, Cohn J, Bonner K *et al.* Scale up of routine viral load testing in  
548 resource-poor settings: current and future implementation challenges. *Clin*  
549 *Infect Dis*. 2016; **62**: 1043–8.
- 550 23. Hermans LE, Moorhouse M, Carmona S *et al* Effect of HIV-1 low-level  
551 viraemia during antiretroviral therapy on treatment outcomes in WHO-  
552 guided South African treatment programmes: a multicentre cohort study.  
553 *Lancet Infect Dis* 2018; **18**:188-97.
- 554 24. Hassan AS, Nabwera HM, Mwaringa SM *et al.* HIV-1 virological failure and  
555 acquired drug resistance among first-line antiretroviral experienced adults at  
556 a rural HIV clinic in coastal Kenya: a cross-sectional study. *AIDS Res Ther*.  
557 2014;**11**:9.



560 **Table 1: Demographic characteristics of the TCEs without genotypes (NG)**  
 561

	M1 Training set	M1 test set	M2 Training set	M2 test set
<b>TCEs</b>	43,239	2,500	50,270	3,000
Patients	13,970	2,500	15,850	3,000
Male	29,290	1,647	33,751	1,947
Female	8,347	544	10,157	704
Not known	5,602	309	6,362	349
Median age	42	41	42	41
<b>Geographical sources of TCEs</b>				
Argentina	112	11	177	27
Australia	481	33	505	34
Brazil	3	1	5	1
Canada	3,554	197	4,214	238
Germany	4,679	244	5,077	266
India	330	37	469	57
Italy	1,418	86	1,632	97
Japan	116	6	133	8
Mexico	308	28	415	34
Netherlands	6,173	368	7,298	464
Romania	434	51	603	68
Serbia	0	0	1	0
South Africa	3,032	303	4,190	451
Spain	4,727	226	5,856	258
UK	9,530	451	10,735	507
USA	1,876	77	2,116	95
Sub-Saharan Africa (country unknown)	52	9	66	10
Unknown (from multinational cohorts/trials)	6,414	372	6,758	385
<b>Baseline data</b>				
Median (IQR) baseline VL ( $\log_{10}$ copies/mL)	3.86 (2.75-4.73)	3.81 (2.65-4.69)	3.83 (2.75-4.7)	3.78 (2.66-4.66)
Median (IQR) baseline CD4 (cells/mm <sup>3</sup> ) (Q1 – Q3) Q2	260 (134-417)	256 (130-405)	261 (139-420)	260 (135-417)
<b>Treatment History</b>				
Median no.(IQR) previous drugs	5 (3 – 8)	4 (3 - 6)	5 (3 - 7)	4 (3 - 6)
N(t)RTI experience (%)	43,119 (99.7)	2,496 (99.8)	50,142 (99.7)	2,995 (99.8)
NNRTI experience (%)	28,653 (66.3)	1,675 (67.0)	33,398 (66.4)	2,026 (67.5)
PI experience (%)	30,030 (69.5)	1,551 (62.0)	34,461 (68.6)	1,776 (59.2)
Median no. (IQR) previous regimens	4 (3 - 9)	4 (2 – 7)	4 (3 - 9)	3 (2 - 7)
<b>New Regimens</b>				
2 N(t)RTI + 1 PI (%)	13,915 (32.2)	861 (34.4)	16,649 (33.1)	1,065 (35.5)
2 N(t)RTI + 1 NNRTI (%)	7,960 (18.4)	550 (22)	9,446 (18.8)	690 (23)
3 N(t)RTIs + 1 PI (%)	2,963 (6.9)	167 (6.7)	3,354 (6.7)	189 (6.3)
3 N(t)RTIs (%)	1,921 (4.4)	94 (3.8)	2,179 (4.3)	111 (3.7)
3 N(t)RTIs + 1 NNRTI (%)	1,493 (3.5)	56 (2.2)	1,661 (3.3)	66 (2.2)
2 N(t)RTIs (%)	1,211 (2.8)	63 (2.5)	1,542 (3.1)	79 (2.6)
2 N(t)RTIs + 1 NNRTI + 1 PI (%)	1,305 (3)	63 (2.5)	1,510 (3)	74 (2.5)
4 N(t)RTIs (%)	750 (1.7)	37 (1.5)	860 (1.7)	43 (1.4)
1 N(t)RTI + 1 NNRTI + 1 PI (%)	883 (2)	43 (1.7)	1019 (2)	49 (1.6)
1 N(t)RTTI + 1 PI (%)	725 (1.7)	42 (1.7)	860 (1.7)	49 (1.6)
Other (%)	10,113 (23.4)	524 (21)	11,190 (22.3)	585 (19.5)

Abbreviations: TCEs (treatment change episodes); IQR (interquartile range); VL (viral load); N(t)RTI (nucleoside or nucleotide reverse transcriptase inhibitor); NNRTI (non-nucleoside reverse transcriptase inhibitor); PI (protease inhibitor)

566  
567

**Table 2: Demographic characteristics of the TCEs including a genotype (G)**

	H1 Training Set	H1 Test Set	H2 Training Set	H2 Test Set
TCEs	18,188	997	17,378	940
Patients	6,844	997	6,700	940
<b>Gender</b>				
Male	11,364	601	10,887	565
Female	2,700	150	2,544	141
Unknown	4,124	246	3,947	234
<b>Geographical sources of TCEs</b>				
Australia	327	24	307	22
Canada	1,773	104	1,616	99
Germany	1,729	84	1,617	74
India	82	6	70	5
International	1,015	51	1,009	51
Italy	869	69	852	66
Japan	115	7	113	7
Netherlands	1,365	74	1297	68
Romania	34	1	34	1
South Africa	176	14	142	12
Spain	2,095	116	2,017	105
Sub-Saharan Africa	44	5	40	4
UK	2,767	118	2,653	114
USA	544	41	521	37
Unknown	5,253	283	5,090	275
<b>Total</b>	<b>18,188</b>	<b>997</b>	<b>17,378</b>	<b>940</b>
<b>Baseline data</b>				
Median (IQR) baseline VL (log <sub>10</sub> c/mL)	4.23 (3.47 - 4.89)	4.3 (3.51 - 4.9)	4.27 (3.5 - 4.9)	4.3 (3.58 - 4.9)
Median (IQR) baseline CD4 (cells/μL)	229 (112 - 380)	240 (113 - 382)	226 (110 - 377)	234 (112 - 380)
<b>Treatment History</b>				
Median (IQR) previous drugs	5 (3 - 8)	5 (3 - 7)	5 (3 - 8)	5 (3 - 7)
N(t)RTI experience (%)	18124 (99.7%)	990 (99.3%)	17317 (99.6%)	933 (99.3%)
NNRTI experience (%)	12019 (66.1%)	641 (64.3%)	11483 (66%)	602 (64%)
PI experience (%)	13015 (71.6%)	717 (71.9%)	12536 (72.1%)	684 (72.8%)
Median (IQR) previous regimens	3 (2 - 7)	3 (2 - 6)	3 (2 - 7)	3 (2 - 6)
<b>New Regimens</b>				
2 N(t)RTI + 1 PI (%)	5887 (32.4)	310 (31.09)	5463 (31.4)	278 (29.6)
2 N(t)RTI + 1 NNRTI (%)	1686 (9.3)	100 (10.03)	1597 (9.2)	97 (10.3)
3 N(t)RTIs + 1 PI (%)	1855 (10.2)	105 (10.53)	1745 (10.0)	99 (10.5)
3 N(t)RTIs (%)	792 (4.4)	32 (3.21)	781 (4.5)	32 (3.4)
3 N(t)RTIs + 1 NNRTI (%)	651 (3.6)	41 (4.11)	634 (3.6)	40 (4.3)
2 N(t)RTIs (%)	490 (2.7)	29 (2.91)	488 (2.8)	28 (3.0)
2 N(t)RTIs + 1 NNRTI + 1 PI (%)	763 (4.2)	50 (5.02)	735 (4.2)	47 (5.0)
1 PI + 1 integrase inhibitor	0 (0)	0 (0)	0 (0)	0 (0)
4 N(t)RTIs (%)	424 (2.3)	17 (1.71)	413 (2.4)	17 (1.8)
1 N(t)RTI + 1 NNRTI + 1 PI (%)	409 (2.2)	38 (3.81)	392 (2.3)	37 (3.9)
1 N(t)RTI + 1 PI (%)	345 (1.9)	14 (1.40)	344 (2.0)	12 (1.3)
Other (%)	4886 (26.9)	261 (26.18)	4786 (27.5)	253 (26.9)

568  
569  
570  
571

Abbreviations: TCEs (treatment change episodes); IQR (interquartile range); VL (viral load); N(t)RTI (nucleoside or nucleotide reverse transcriptase inhibitor); NNRTI (non-nucleoside reverse transcriptase inhibitor); PI (protease inhibitor)

572

**Table 3: Results of the modeling without a genotype (NG)****A: Cross validation during model development**

NG1 models (standard data windows)						NG2 models (expanded data windows)				
		Sens	Spec				Sens	Spec	OA	
Model	AUC	(%)	(%)	OA (%)	oop	AUC	(%)	(%)	(%)	oop
1	0.84	71%	81%	78%	0.43	0.84	71%	80%	77%	0.42
2	0.84	71%	80%	77%	0.42	0.83	71%	79%	76%	0.42
3	0.83	71%	80%	77%	0.42	0.83	71%	79%	76%	0.41
4	0.83	71%	80%	77%	0.42	0.83	71%	79%	76%	0.42
5	0.83	71%	80%	77%	0.43	0.83	71%	79%	76%	0.43
avg	0.84	71%	80%	77%	0.42	0.83	71%	80%	76%	0.42
min	0.83	71%	80%	77%	0.42	0.83	71%	79%	76%	0.41
max	0.84	71%	81%	78%	0.43	0.84	71%	80%	77%	0.43
B: Independent testing										
Main test set										
NG1 n=2,500	0.82	72%	77%	75%	0.42	0.81	73%	76%	75%	0.42
NG2 n=3,000										
Marginal cases*										
(N= 500)						0.79	73%	73%	73%	0.42
New drug subsets										
EVG (50)	0.81	80 %	79%	80%	0.61	0.75	74%	79%	76%	0.63
MVC (50)	0.89	88 %	89%	88%	0.52	0.83	74 9%	74%	74%	0.52
TPV (50)	0.84	71%	73%	73%	0.35	0.85	70%	82 %	80%	0.43

OOP: optimum operating point OA: overall accuracy

\* Cases with baseline data outside the standard windows used for M1

Table 4: Results of modeling with a genotype (G)

A: Cross validation during model development

Model	G1 models (standard non-adherence filter)					G2 models (experimental filter)				
	AUC	Sens	Spec	OA	oop	AUC	Sens	Spec	OA	oop
1	0.86	71%	78%	78%	0.41	0.86	78%	79%	79%	0.43
2	0.86	73%	82%	79%	0.43	0.89	81%	82%	82%	0.42
3	0.86	73%	82%	79%	0.43	0.89	81%	81%	81%	0.42
4	0.84	73%	80%	78%	0.42	0.87	77%	80%	79%	0.41
5	0.86	73%	82%	79%	0.43	0.88	79%	81%	80%	0.41
avg	0.86	73%	81%	79%	0.42	0.88	79%	81%	80%	0.42
min	0.84	71%	78%	78%	0.41	0.86	77%	79%	79%	0.41
max	0.86	73%	82%	79%	0.43	0.89	81%	82%	82%	0.43
B: Independent testing										
Test set	AUC	Sens	Spec	OA	oop	AUC	Sens	Spec	OA	oop
G1 n=997	0.84	72%	80%	76%	0.42	0.83	75%	76%	76%	0.42
G2 n=940	0.85	72%	82%	77%	0.42	0.86	74%	83%	79%	0.42

OOP: optimum operating point      OA: overall accuracy

Table 5: Comparison of model predictions versus GSS for test TCEs with genotypes

Prediction system	AUC	Sensitivity (%)	Specificity (%)	Overall accuracy (%)	p (GSS versus either models)
NG1 models	0.81	71	77	75	
NG2 models	0.81	68	78	74	
Total ANRS score	0.56	53	55	54	<0.0001
Total HIVdb score	0.58	40	68	56	<0.0001
Total REGA score	0.55	53	53	53	<0.0001
G1 models	0.84	73	80	77	
G2 Models	0.86	76	82	80	
Total ANRS score	0.61	53	61	58	<0.0001
Total HIVdb score	0.63	54	63	59	<0.0001
Total REGA score	0.60	57	57	57	<0.0001

GSS: genotype sensitivity score



593 **Table 6: Results of in silico analysis for the G and NG models**

<b>A: NG models</b>			
<b>Test set</b>	<b>Measure</b>	<b>NG1 models</b>	<b>NG2 models</b>
All cases with $\geq 3$ drugs in their new regimen	Alternatives predicted to be effective (%)	97	98
M1 = 2,315 cases and M2 = 2,783 cases*	Alternatives with higher probability of response than regimen used in clinic	100	100
Subset of the above that failed the new regimen introduced in the clinic	Alternatives Predicted to be effective (%)	96	96
M1 = 1,433 cases and M2 = 1716 cases*	Alternatives with higher probability of response than regimen used in clinic	100	100
<b>B: G models</b>		<b>G1</b>	<b>G2</b>
All cases with $\geq 3$ drugs in their new regimen	Alternatives predicted to be effective (%)	93	93
(H1 = 892, H2 = 841 cases*)	Alternatives with higher probability of response than regimen used in clinic (%)	99.9	99.9
Subset of the above that failed the new regimen introduced in the clinic	Alternatives Predicted to be effective (%)	90	90
(H1 = 564 cases and H2 = 513 cases*)	Alternatives with higher probability of response than regimen used in clinic (%)	100	100

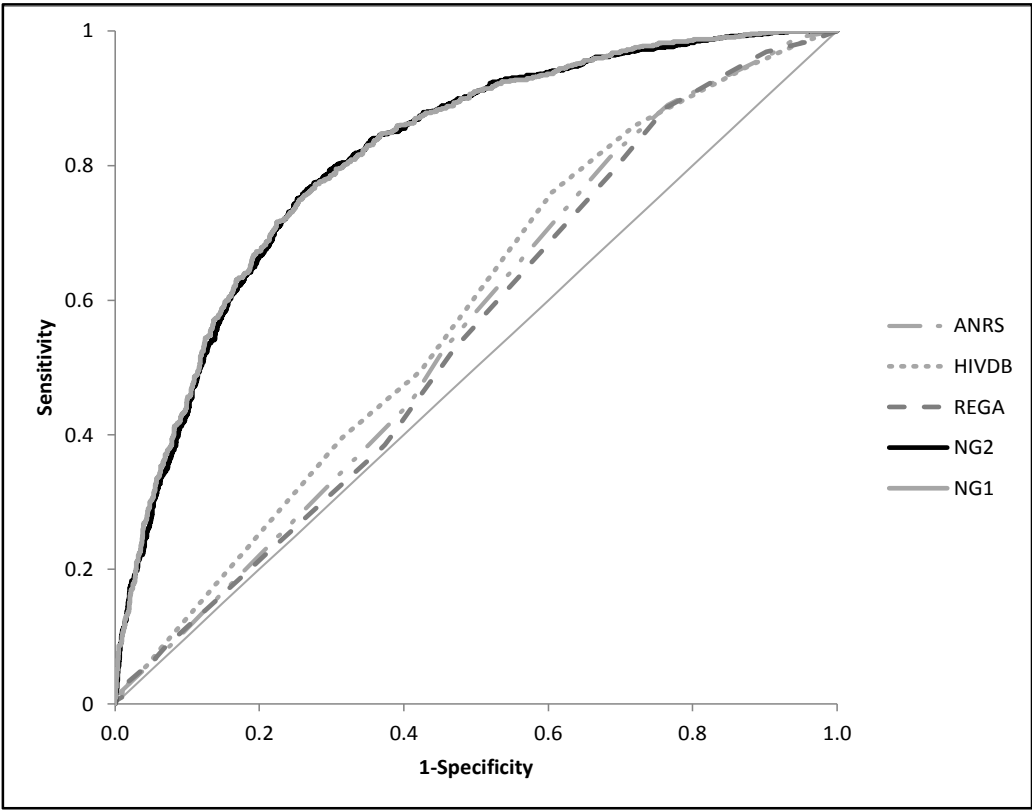
594 \* only those cases with  $\geq 3$  drugs in the new regimen were used in these analyses

595

596

597

598 **Figure 2: ROC curves for the NG models versus genotyping (GSS)**  
599  
600



601  
602  
603  
604

**Figure 3: ROC curves for the G models versus genotyping (GSS)**

