

# Improving the prediction of virologic response to tipranavir: the development of a tipranavir weighted score

J Scherer<sup>1</sup>, CA Boucher<sup>2</sup>, JD Baxter<sup>3</sup>, JM Schapiro<sup>4</sup>, VM Kohlbrenner<sup>1</sup>, DB Hall<sup>1</sup>

<sup>1</sup> Boehringer Ingelheim Pharmaceuticals, Ridgefield, CT, USA;

<sup>2</sup> Department of Virology, University Medical Center Utrecht, the Netherlands;

<sup>3</sup> Cooper Hospital/UMDNJ-Robert Wood Johnson Medical School, New Jersey, USA;

<sup>4</sup> National Hemophilia Center, Israel

## Introduction

The methodologies used to develop mutation scores to predict susceptibility to individual antiretroviral agents (ARVs) is in constant revision and thus it is not surprising, that when new methodologies are also applied to the original tipranavir score [1] the impact of the individual mutations identified are not the same, and appropriate adjustments for the contribution of each mutation towards virologic response are necessary. Also, it has been observed that several mutations result in increased response or reduced likelihood of virologic failure and these mutations also need to be considered in the development of a mutation score to determine as accurately as possible the level of activity of a specific ARV. The purpose of this work was to develop weights for each of these mutations to arrive at a weighted TPV mutation score that will provide guidance to treating physicians on the impact of individual protease mutations on TPV activity.

## Methods

### Datasets

The base dataset consisted of all RESIST patients who had baseline genotype, were randomized and received at least one dose of tipranavir resulting in N=745 patients. This base dataset was randomly broken up into a score development dataset and an evaluation dataset consisting of roughly 75% and 25% of the base dataset (N=566 and N=179 for the development and the test datasets, respectively). The evaluation dataset was used for assessing the predictive accuracy of the mutation score and comparing it to other commonly used scores.

The RESIST study population was specific by design in excluding patients whose baseline genotype consisted of  $\geq 3$  substitutions at codons 33, 82, 84 and 90. Therefore it should be mentioned that some mutational patterns were not present in the dataset by design of the studies.

### Score development

1. **Mutations** The initial list of mutations for consideration was restricted to the mutations in the existing TPV mutation score plus 5 mutations believed to increase susceptibility to TPV:

TPV score mutations – 10V, 13V, 20M/R/V, 33F, 35G, 36I, 43T, 46L, 47V, 54A/M/V, 58E, 69K, 74P, 82L/T, 83D, 84V [1].

Mutations associated with observed increased susceptibility to TPV – 24I, 30N, 50L/V, 54L, 76V.

2. **Models** Six models relating the mutations above to different response variables were run (Table 1).

Table 1. Models run relating mutations to virologic response

Model	Week	Variable	Method+	Analysis*
1	8	Change from baseline in VL	OT	ANOVA
2	24	Change from baseline in VL	LOCF	ANOVA
3	8	At least 1 log drop in VL	NCF	Logistic regression
4	24	At least 1 log drop in VL	NCF	Logistic regression
5	8	VL <400	NCF	Logistic regression
6	24	VL <400	NCF	Logistic regression

\* All models adjusted for baseline CD4+ cell counts and OBR activity.  
OT = On-Treatment – patients with missing values not included in analyses; LOCF = Last On-treatment Carried Forward; NCF = Non-Completers considered Failures.

3. **Cross-validation** The test dataset was randomly divided into 10 subsets and each model was run 10 times, each time leaving out one of the subsets. This allowed both an assessment of the robustness of the model estimates (and hence weights) and also an independent, cross-validation estimate of the association of the weighted score with virologic response, after adjusting for baseline CD4+ cell count and background activity.

4. **Determination of weights** For each model (6 models) and run (10 runs), the weights for each mutation were formed as follows (note that for the logistic regression models, the log (odds ratios) were used rather than the parameter estimates but the methodology was the same):

- The mutation with the largest absolute model estimate (max estimate = M) was given a weight of 10 (if "harms" virologic response) and -10 (if "helps" virologic response).
- The other weights (W) were computed to keep the relationship of the parameter estimates (E) intact, with the integer value being retained ( $W = \text{floor}(10 \cdot (E / M))$ ).

Table 2 shows the model estimates and odds ratios for run 3 (i.e. excluding the 10% of patients randomly selected into subset 3) and the corresponding calculation of weights for each model.

Table 2. Summary of model estimates (odds ratios) and resulting weights of each model for run 3 out of 10

Mutation	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Est	Weight	Est	Weight	OR	Weight	OR	Weight	OR	Weight	OR	Weight
10V	0.07	1	0.27	4	1.13	-1	0.75	2	0.99	0	0.58	2
13V	0.02	0	0.04	0	0.74	3	0.79	1	0.87	1	0.74	1
20M/R/V	-0.03	0	-0.12	-2	0.86	1	0.99	0	1.03	0	0.99	0
24I	-0.36	-7	-0.13	-2	2.65	-10	1.39	-2	1.14	0	1.13	0
30N	-0.11	-1	-0.09	0	1.53	-4	0.64	3	0.94	0	0.54	2
33F	0.03	0	0.06	0	0.89	1	0.90	0	1.32	-2	1.27	0
35G	0.19	3	-0.52	-8	1.05	0	0.57	4	0.77	1	1.03	0
36I	0.12	2	0.35	5	0.75	2	0.47	6	0.59	3	0.56	2
43T	0.10	2	0.22	3	0.73	3	0.82	1	1.00	0	0.60	2
46L	0.18	3	0.03	0	0.65	4	0.74	2	0.81	1	0.68	1
47V	0.34	6	0.29	4	0.44	8	0.48	6	0.26	10	0.15	7
50L/V	-0.47	-9	-0.10	-1	2.23	-8	0.91	0	1.04	0	0.80	0
54A/M/V	0.21	4	0.23	3	0.62	4	0.77	2	0.67	2	0.67	1
54L	-0.37	-7	-0.54	-8	1.04	0	3.19	-9	3.65	-9	6.34	-7
58E	0.32	6	0.35	5	0.59	5	0.48	6	0.64	3	0.43	3
69K	0.12	2	-0.10	-1	0.77	2	1.02	0	0.74	2	0.99	0
74P	0.49	10	0.62	10	1.24	-2	0.29	10	0.40	6	0.47	2
76V	-0.31	-6	-0.21	-3	1.82	-6	1.56	-3	1.30	-1	1.22	0
82L/T	0.37	7	0.26	4	0.41	9	0.81	1	0.72	2	0.72	1
83D	0.32	6	0.43	7	0.60	5	0.40	7	0.54	4	0.08	10
84V	0.15	3	0.18	2	0.71	3	0.77	2	0.82	1	0.55	2
R-s*	0.48		0.48		0.46		0.46		0.49		0.50	
Wscore P*	0.0054		0.0044		0.0142		0.0156		0.0031		0.0018	

Est = Adjusted mean response without mutation – adjusted mean response with mutation (i.e. estimated amount of "harm" having the mutation will have on the virologic response – negative value implies the mutation is "helping" the response to drug).

OR = odds response / odds non-response.

\* From regressing week 8 change in VL on baseline CD4+ cell count, background activity and the resulting weighted score using the N=70 patients contained in cross-validation subset 3. Wscore P is the p-value for showing the statistical significance of the weighted score.

5. **Final weights** Weights for each of the 6 models were determined by taking the median weight for the runs with the best fit on the cross-validation datasets. The final weight was determined by taking the median weight across the six models. The derivation of the final weights is shown in Table 3.

**Determining clinical cut-offs** The cut-offs for predicting a patient to be susceptible (S), partially susceptible (PS) or resistant (R) to TPV were determined based on the Area Under the Receiver Operator Curve (AUROC) – the cut-points were chosen that resulted in the largest AUROC, after adjusting for baseline CD4+ cell count and background antiviral drug activity, a measure of prediction accuracy. The entire dataset of TPV/r patients in RESIST was used to determine the cut-offs.

### Comparison to other scores

Roughly a quarter of the base dataset (N=179) was left out of the score development to provide an independent source to compare to other available mutation scores.

**Comparing to other scores** The tipranavir weighted score (TW) was compared to the tipranavir unweighted score [1] (TUW), the score from the Stanford website [2] (STAN), the score from Rega Institute for Medical Research version 7.0® (REGA) [3], and the score from the Virtual Phenotype® (VP) methodology from Virco. Comparisons were made at weeks 8 and 24 by looking at:

- Spearman correlation coefficients of weeks 8 (On-Treatment – missing data not replaced) and 24 (Missing values replaced by carrying forward the last observation on treatment) viral load changes from baseline.
- Logistic regression models were run for each score with weeks 8 (at least a 1 log drop in viral load) and 24 (VL <50) virologic response as the dependent variable and predictors of TPV mutation score and OBR activity. The comparisons made between scores were the p-value assessing statistical significance of each mutations score and AUROC to compare overall prediction accuracies.

### Background Activity Scores

The development of the weighted score and (model-based) comparisons to other scores were all adjusted for the activity of the background regimen. The estimated activities for each patient were based on a novel approach developed by Hall et al. [4] that estimates the contribution of each drug in the Optimized Background Regimen using linear model-based approach accounting for the baseline resistance (using Virtual Phenotype®) and historical use of each drug. Often, the activity of the drug in the background regimen, especially with the NRTIs, is largely dependent on the previous pattern of use (due to archived resistance) rather than the baseline resistance. These adjustments prove to be much better at predicting response than a straight count of the number of active drugs in the OBR.

## Results

### Model fitting and determination of weights

The final weights are shown in Table 3 and can be divided in three categories as shown in Table 4. Some notes about the final weights:

- 33F, previously believed to be a key tipranavir mutation, receives a weight of 0 and thus has been removed from the weighted score.
- 84V, a major mutation for all PIs (receives a weight of 25 in the Stanford score), receives only a weight of 2 in the revised analyses.
- Mutations 54L, 50V and 76V, all representing major mutations contributing resistance to fosamprenavir and also to darunavir, predict improved response to tipranavir.
- Only 47V (14%), 54A/M/V (64%) and 58E (16%) were major mutations seen in over 10% of patients from the tipranavir development program. This is the reason such a small number of patients entering the RESIST trials were resistant to tipranavir providing further evidence of the unique resistance profile of tipranavir compared to the other protease inhibitors.
- TPV score mutations 13V and 69K, which were seen more frequently in non-clade B patients in the BI-sponsored trial of naive patients 1182.33 (13V: 17.2% B vs 48.4% non-B and 69K: 4.0% B vs 66.2% non-B), both receive weights of 0 and are thus removed from the weighted score.

## Poster Number P3.4/07

11th European AIDS Conference (EACS),  
Madrid, Spain  
24-27 October 2007

Corresponding Author: Joseph Scherer  
Boehringer Ingelheim Pharmaceuticals,  
Ridgefield, CT, USA  
Tel: +49 (6132) 77-97217  
Fax: +49 (6132) 72-97217  
Email: Joseph.Scherer@ing.boehringer-ingenheim.com

Table 3. Summary of combined weights for each model

Mutation	Model <sup>1</sup>						Final weight <sup>2</sup>	Incidence <sup>3</sup>
	1	2	3	4	5	6		
10V	1	3	-1	2	0	1	1	12.9
13V	0	0	3	2	0	1	0	34.8
20M/R/V	0	-2	1	0	0	0	0	34.4
24I	-6	-2	-8	-2	0	0	-2	15.5
30N	-2	-1	-3	1	0	1	0	3.8
33F	1	0	1	0	-2	0	0	28.7
35G	3	-7	0	4	0	0	0	1.0
36I	1	4	2	6	3	2	2	52.2
43T	2	3	3	1	0	2	2	13.6
46L	2	0	3	2	0	1	1	20.8
47V	6	4	8	6	10	7	6	14.0
50L/V	-8	-1	-7	0	0	0	-4	7.1
54A/M/V	4	4	4	2	2	2	3	63.7
54L	-8	-7	0	-7	-9	-7	-7	7.1
58E	6	5	5	6	2	3	5	15.7
69K	2	-2	2	-1	0	-1	0	5.1
74P	8	10	7	4	5	6	6	3.4
76V	-3	-2	-4	-3	-1	0	-2	8.4
82L/T	7	7	10	3	2	1	5	4.7
83D	1	7	2	8	1	10	4	1.4
84V	3	2	3	2	1	2	2	30.1

Runs<sup>1</sup> 2,3,4,5 2,3,4,5 2,3,4,5 2,3,4,5 2,3,5 2,3,4,6

<sup>1</sup> Weights for each model are arrived at by taking the median (rounded down) of all runs with significant p-values for the resulting weighted score when regressed against week 8 VL change from baseline, using the cross-validation datasets. The bottom row of the table shows which runs were used in the calculation of the median.

<sup>2</sup> Median weight across models (also rounded down).

<sup>3</sup> % of baseline isolates containing that mutation from the phase III TPV development program (excluding naive trial 1182.33).

<sup>4</sup> Weight taken as the average of models 1 and 2 due to inconsistency of results and strong correlation with 50L/V and TPV phenotype.

Table 4. Weight classification

Class	Weight	Mutations
Increased response	<0	24I, 50L/V, 54L, 76V
Minor mutations	-1-2	10V, 36I, 43T, 46L, 84V
Major mutations	>2	47V, 54A/M/V, 58E, 74P, 82L/T, 83D

### Performance and comparison to other scores

Cut-offs for the weighted score were determined using all RESIST TPV/r-treated patients. A lower cut-off of 3 and upper cut-off of 10 yielded the largest AUROC (78.5%) for predicting 24 week VL <50 copies/mL, after adjusting for baseline CD4+ cell count and OBR activity. The score levels with frequencies and numbers of responders are provided in Table 5. Also included is a summary of the response rates at each cut-point.

An increasing number of major mutations (47V, 54A/M/V, 58E, 74P, 82L/T, 83D) proves to be a good marker for the resulting resistance classifications. A cross-classification of number of major mutations with clinical interpretation of the weighted score is included in Table 6. These data suggest the following regarding the interpretation of the tipranavir weighted score:

- You need at least 2 major mutations to be resistant to tipranavir and with  $\geq 3$  full resistance was seen in all isolates available.
- With  $\leq 1$  of the major mutations, TPV/r is likely to retain partial susceptibility and with no major mutations nearly all isolates retained full susceptibility to TPV/r.

The results of the comparisons to other scores using the independent dataset are provided in Table 7. The TW and VP both predict better than the other three scores at both weeks 8 and 24, with VP having the highest correlation with viral load decline ( $R = 0.25$ ) and prediction accuracy (71.0%) at week 8 and TW being the best at week 24 ( $R = 0.35$  and AUROC = 71.2%). It makes sense that these two scores would perform the best since they are both based on a similar methodology, that being a linear-model based approach to estimate the effects of each mutation. Slight differences are that VP is developed on a much larger database and is designed to predict phenotype, whereas TW is designed to predict virologic response.

Table 5. Frequencies and virologic response by weighted score using all RESIST TPV/r patients

Score	N	Number of responders		
		Week 8	Week 24	Week 48
Overall	745	419(56.2%)	169(22.7%)	169(22.7%)
-7	2	2	0	0
-6	3	2	1	0
-5	3	3	2	2
-4	9	9	6	6
-3	5	5	1	2
-2	9	7	2	4
-1	10	7	0	2
0	66	38	24	26
1	43	32	16	16
2	63	47	21	18
3	68	45	20	20
Partially Susceptible (PS): $\leq 3$	283	197(69.6%)	93(32.9%)	96(33.9%)
4	51	31	13	15
5	96	47	15	13
6	57	30	11	11
7	66	31	8	9
8	35	18	8	7
9	34	14	6	6
10	28	13	9	7
Resistant (R): $> 10$	95	38(40.0%)	6(6.3%)	5(5.3%)

Table 6. Cross-classification of clinical interpretation using TW with number of major mutations

Number of patients	N (%) of major mutations			
	0	1	2	3
Interpretation	184	374	160	27
Susceptible	168(91.3%)	115(30.8%)	0(0.0%)	0(0.0%)
Partially susceptible	16(8.7%)	259(69.3%)	92(57.5%)	0(0.0%)
Resistant	0(0.0%)	0(0.0%)	68(42.5%)	27(100.0%)

Table 7. Comparisons with other commonly used tipranavir mutation scores on the independent dataset

Score	R	Week 8		R	Week 24	
		P-value	AUROC		P-value	AUROC
TW	0.23	0.0235	70.1%	0.35	0.1048	71.2%
TUW	0.11	0.4916	68.5%	0.25	0.0875	69.5%