Comparison of tests and procedures to build clinically relevant genotypic scores: application to the Jaguar study

Philippe Flandre1*, Anne-Geneviève Marcelin2, Juliette Pavie3, Nathalie Shmidely4, Marc Wirden2, Olivier Lada2, Marie-Charlotte Bernard4, Jean-Michel Molina4 and Vincent Calvez2

1INSERM U 472, Paul Brousse Hospital, Villejuif, France
2Department of Virology, Pitié-Salpêtrière Hospital, Paris, France
3Department of Infectious Diseases, Saint Louis Hospital, Paris, France
4Bristol-Myers-Squibb France Laboratory, Rueil-Malmaison, France

*Corresponding author: Tel: +33 1 455 95071; Fax: +33 1 455 95169; E-mail: flandre@vjf.inserm.fr

Antiviral Therapy 10:479–487

© 2005 International Medical Press 1359-6535

Objective: To compare non-parametric tests and procedures of selection in building clinically validated genotypic scores.

Design and patients: In the Jaguar study, 111 patients on a stable antiretroviral regimen experiencing virological failure were randomized in the didanosine (ddI) arm to receive ddI for 4 weeks in addition to their current combination therapy.

Methods: The virological response was HIV-1 RNA reduction from baseline to week 4. The univariate impact of each mutation associated with resistance to ddI on virological response was quantified by comparing reduction in plasma HIV-1 RNA in patients with or without the specific mutation, using a Wilcoxon–Mann–Whitney test. The next step was to select the combination of mutations most strongly associated with the virological response. Two procedures and two tests were compared using either the set of resistance mutations or the set of resistance mutations and mutations providing a better virological response. The Kruskal–Wallis and the Jonckheere test for ordered alternatives were compared in order to build a genotypic score using the two distinct procedures.

Results: Eight mutations were associated with a reduced virological response to ddI: M41L, D67N, T69D, L74V, V118I, L210W, T215Y/F and K219Q/E and two mutations with a better virological response: K70R and M184V/I. The Jonckheere–Terpstra test for trend provided the combination of mutations (M41L+T69D-K70R+L74V-M184V/I+T215Y/F+K219A/E) that were the most predictive for the week 4 virological response, that is, leading to the lowest P value. The ‘removing’ procedure, starting from a set of mutations retained and removing mutations one by one to find the best combination, provides lower P values than the ‘adding’ procedure starting with a single mutation and adding mutations one by one. Whatever the set of mutations and the procedure used, the Jonckheere–Terpstra test selects combinations of mutations leading to lower P values than the Kruskal–Wallis test.

Conclusion: The Jonckheere–Terpstra test for trend is recommended for building a genotypic score when compared with the Kruskal–Wallis. The choice of the selection procedure is discussed here and may be dependent on the objective of the score.

Introduction

There has been increasing acceptance of resistance testing as an important tool in making decisions about antiretroviral therapy. In particular, several studies have shown that genotypic antiretroviral resistance testing results in a better short-term virological outcome than the standard of care in patients who are about to start their next antiretroviral regimen [1–5]. Now that virological benefit has been established for selecting antiviral drugs based on baseline resistance mutations, attention is being focussed on ways of determining clinically relevant interpretation rules or algorithms [6–9].

The difficulty in building these rules includes the presence of undetected/archived/minor species, complexity and variability of resistant variants, interactions between mutations and lack of standardized statistical methods. A major concern in this field is the choice of the statistical methodology. Different statistical methods have been used including non-parametric tests, regression modelling, discriminant analysis, recursive partitioning and artificial neural networks [10–15].

A simple approach, however, that has been widely used consists of defining a genotypic score as a simple
The combination of mutations defining the score is selected, by a non-parametric test, as the one providing the strongest association with virological response. Combinations of mutations are investigated from a list of mutations selected by a univariate approach, comparing patients with and without the corresponding mutation. This work discusses the different statistical procedures that can be used to provide such genotypic scores. The two key components of such analyses are the non-parametric test used and the procedure of selection. The Kruskal–Wallis non-parametric test is widely used but is not well adapted to this setting and the Jonckheere–Tepstra ([JT] also known as the Jonckheere) test should be used in preference [17]. We also suggest two procedures to select the combination of mutations providing the strongest association with the virological outcome. A recent clinical trial is used to illustrate the different statistical procedures [18].

Patients and methods

Patients and virological endpoint

The Jaguar study was a randomized, multicentre, double-blind, placebo-controlled trial, evaluating the efficacy of adding ddI to an ongoing, failing highly active antiretroviral therapy (HAART) regimen [18]. The only component then added to the failing regimen was ddI or ddI-placebo and the randomization ensured that the other drugs were well balanced between the two groups. Subjects were recruited from 29 clinical centres located throughout France and must have been treated for at least 3 months with a stable antiretroviral drug regimen that did not include ddI, although ddI could have been part of treatment previous to that.

The virological endpoint was the viral load reduction between baseline and week 4. This endpoint is usually the most informative. The difficulty arises, however, when patients reach undetectable HIV-1 RNA levels due to the limit of quantification of the virological assay [19–21]. In the ddI arm, only 11 out of 102 (11%) patients reached the limit of quantification of 50 copies/ml indicating that the naive method can be used, that is, all HIV-1 RNA measurements of <50 copies/ml were fixed at 50 copies/ml.

Statistical methods

Selection procedures

Whatever the procedure and non-parametric test used, the first step was to select a set of mutations by a univariate analysis. This analysis compares, for each position, the virological response obtained with and without the presence of the variant. From the IAS-USA list, mutations present in at least 5% of the samples and leading to univariate $P$ value lower than 0.20 can be considered as eligible for the multivariate analysis [22]. One can consider only mutations providing poorer virological response or also mutations associated with a better response, that is, mutations increasing the drug susceptibility. In the first situation, mutations are simply added in a variable $Z$ and each mutation has an equal weight of +1. For example, $Z$ = M41L+D67N+T215Y/F where each position 41, 67 and 215 of the corresponding variable takes the value of 1 if the mutation is present and 0 if absent ($Z$ varying from 0–3). In the second case, mutations providing a better response had a weight of –1. For example, it is known that M184V/I may re-sensitize TAM-resistant virus to specific regimens containing zidovudine or stavudine. In this situation, we can have $Z$ = M41L+D67N–M184VI, where $Z$ varies from –1 (only M184V/I is present) to 2 (M184V/I is absent and both M41L and D67N are present). Using the non-parametric tests described below, groups of patients to be compared are defined by distinct values of $Z$. For ease of presentation, the procedure is described for the first situation where mutations are simply added.

When mutations are eligible from the univariate analysis, either all combinations of mutations are investigated or a step-by-step procedure is used to select the final set of mutations most strongly associated with virological response. The first case leads to a comparison of a great number of combinations involving a multiple comparison approach without any selection procedure. For instance, with eight mutations eligible, 246 different combinations of two or more mutations would be compared (246 = 8!/2!6! + 8!/3!5! + 8!/4!4! + 8!/5!3! + 8!/6!2! + 8!/7!1! + 8!/8!). We suggest instead a second possibility, comparing two different procedures to select the final set of mutations. The ‘adding’ procedure consists of addition of mutations one by one to find the best combination of mutations associated with the virological response. The first mutation retained is the mutation associated with a better response, that is, mutations associated with a poorer virological response or also mutations associated with a better response had a weight of +1. For example, it is known that M184V/I may re-sensitize TAM-resistant virus to specific regimens containing zidovudine or stavudine. In this situation, we can have $Z$ = M41L+D67N–M184VI, where $Z$ varies from –1 (only M184V/I is present) to 2 (M184V/I is absent and both M41L and D67N are present). Using the non-parametric tests described below, groups of patients to be compared are defined by distinct values of $Z$. For ease of presentation, the procedure is described for the first situation where mutations are simply added.

The virological endpoint was the viral load reduction between baseline and week 4. This endpoint is usually the most informative. The difficulty arises, however, when patients reach undetectable HIV-1 RNA levels due to the limit of quantification of the virological assay [19–21]. In the ddI arm, only 11 out of 102 (11%) patients reached the limit of quantification of 50 copies/ml indicating that the naive method can be used, that is, all HIV-1 RNA measurements of <50 copies/ml were fixed at 50 copies/ml.

Statistical methods

Selection procedures

Whatever the procedure and non-parametric test used, the first step was to select a set of mutations by a univariate analysis. This analysis compares, for each position, the virological response obtained with and without the presence of the variant. From the IAS-USA list, mutations present in at least 5% of the samples and leading to univariate $P$ value lower than 0.20 can be considered as eligible for the multivariate analysis [22]. One can consider only mutations providing poorer virological response or also mutations associated with a better response, that is, mutations increasing the drug susceptibility. In the first situation, mutations are simply added in a variable $Z$ and each mutation has an equal weight of +1. For example, $Z$ = M41L+D67N+T215Y/F where each position 41, 67 and 215 of the corresponding variable takes the value of 1 if the mutation is present and 0 if absent ($Z$ varying from 0–3). In the second case, mutations providing a better response had a weight of –1. For example, it is known that M184V/I may re-sensitize TAM-resistant virus to specific regimens containing zidovudine or stavudine. In this situation, we can have $Z$ = M41L+D67N–M184VI, where $Z$ varies from –1 (only M184V/I is present) to 2 (M184V/I is absent and both M41L and D67N are present). Using the non-parametric tests described below, groups of patients to be compared are defined by distinct values of $Z$. For ease of presentation, the procedure is described for the first situation where mutations are simply added.

When mutations are eligible from the univariate analysis, either all combinations of mutations are investigated or a step-by-step procedure is used to select the final set of mutations most strongly associated with virological response. The first case leads to a comparison of a great number of combinations involving a multiple comparison approach without any selection procedure. For instance, with eight mutations eligible, 246 different combinations of two or more mutations would be compared (246 = 8!/2!6! + 8!/3!5! + 8!/4!4! + 8!/5!3! + 8!/6!2! + 8!/7!1! + 8!/8!). We suggest instead a second possibility, comparing two different procedures to select the final set of mutations. The ‘adding’ procedure consists of addition of mutations one by one to find the best combination of mutations associated with the virological response. The first mutation retained is the mutation providing the lower $P$ value (higher association with the virological response), then all possible combinations of two mutations are compared through the virological response of patients harbouring none, one and two mutations of the combination investigated. The combination providing the lower $P$ value is retained and the procedure continues with all possible combinations of the three mutations based on the first two mutations selected. The procedure stops when addition of a new mutation does not provide a lower $P$ value.

The ‘removing’ procedure consists of removing mutations one by one, starting with the K mutations retained from the first analysis. From the initial set of K mutations, all mutations are removed one by one to investigate all combinations of K–1 mutations. Again the non-parametric test will compare groups of patients...
having none to \(K-1\) mutations and the combination of \(K-1\) mutations providing the lower \(P\) value is retained. In the second step, mutations are again removed one by one to compare the different combinations of \(K-2\) mutations, the combination providing the lower \(P\) value is again retained, and so on. The procedure stops when removing a mutation does not provide a lower \(P\) value than the previous \(P\) value.

**Non-parametric tests**

The Wilcoxon–Mann–Whitney (WMW) test is well adapted for univariate analysis to select the mutations eligible for the genotypic score by comparing groups of patients with and without the mutation. For the two procedures, we compared the use of two non-parametric tests to select the final combination of mutations.

The natural extension of more than two independent samples of the WMW test is the Kruskal–Wallis (KW) test. The KW statistic tests the null hypothesis that \(n\) samples come from the same population or from identical populations with the same median [17]. To specify the null hypothesis and its alternative more explicitly, let \(\theta\) be the population median for the \(j\)th group where \(j\) represents the number of mutations. Then we may write the null hypothesis that the medians are the same as \(H_0: \theta_0 = \theta_1 = \ldots = \theta_n\) and the alternative hypothesis may be written as \(H_1: \theta_i \neq \theta_j\) for some groups \(i\) and \(j\). Therefore, if the alternative hypothesis is true, at least one pair of groups has different medians. As described by the alternative hypothesis, the KW test is well adapted to compare nominal groups, for example, comparing virological response between patients having distinct modes of transmission, but it is less appropriate to compare ordinal groups. Indeed, when comparing groups of patients with distinct numbers of resistance mutations one would expect that patients with no resistance mutations would have a better virological response than patients with one mutation, who in turn would have a better response than patients with two mutations, and so on. When mutations provide a better response, the above assumption is true whether a negative weight is attributed to such mutations, as described previously.

The JT test, for ordered alternatives has been specially designed for such a context [23]. The null hypothesis of the JT test can be written as above, while the alternative hypothesis is \(H_1: \theta_0 \geq \theta_1 \geq \ldots \geq \theta_n\), that is the medians are ordered in magnitude. It is important to note that in order to ensure proper use of the test, the order of the groups should be determined \textit{a priori}, which is obvious in our setting. It is clear that the JT test is more powerful and more appropriate than the KW test, which, although valid, is too general.

**Results**

Inadequacy of the KW test is illustrated with hypothetical data in Figure 1. Data in Figure 1A and B are exactly the same but groups with 0, 1, 2, 3 and 4 mutations in Figure 1A become groups with 1, 3, 0, 4 and 2 mutations, respectively, in Figure 1B. The unspecific KW test provides exactly the same \(P\) value \((P=0.023)\) for the two plots while the JT test leads to \(P=0.0004\) when ordered virological response corresponds to ordered number of mutations, and \(P=0.187\) when this is not the case. Although the JT is clearly the most powerful test in our context, we will compare results with the KW test as this test has been extensively used.

From the ddI group, 102 patients had both HIV-1 RNA reduction at week 4 and genotype available. Both naive and censored methods provide a median decrease of \(0.57 \log_{10} \text{copies/ml} \) with interquartile ranges (IQR) of \(0.14-1.02\) and \(0.15-1.03\), respectively. Table 1 summarizes results of the univariate analysis with a Wilcoxon test for mutations that were present in at

![Figure 1. Virological response on hypothetical data to compare the KW test and the JT test for trend](image-url)
least 5% of the sample and with a corresponding
$P < 0.20$. In Table 1, the first eight mutations listed,
which are associated with a poorer virological
response, are ordered from lower to higher
$P$ values while the last two mutations (K70R and M184VI) are
associated with a better virological response.

KW test
Firstly, only the eight mutations (set I: M41L, D67N,
T69D, L74V, V118I, L210W, T215Y/F and K219Q/E),
associated with a poorer virological response,
were eligible for the selection procedures. All possible combi-
nation of two mutations were analysed and they did not
provide lower $P$ values than M41L alone. Indeed, the
exact $P$ value associated with position 41 is $1.1 \times 10^{-5}$
while the lower $P$ value associated with a combination
of two mutations (M41L+L74V) is $1.45 \times 10^{-5}$. Then,
using this procedure, the genotypic ‘score’ is defined
only from presence or absence of the M41L mutation.
Using the ‘removing’ procedure leads to the selection of the combination
of M41L and L74V, with a corresponding score taking the value –1, 0 or 1 and associated $P$ value of $2.11 \times 10^{-6}$. The ‘removing’ procedure selects
M41L+D67N+L184VI+L210W+K219Q/E to define
the genotypic score with a corresponding $P$ value of
$2.02 \times 10^{-7}$. HIV-1 RNA reduction for the two genotypic
scores provided from both original sets I and II using
the ‘removing’ procedure are displayed in Figure 2.
This figure provides an illustration with genuine clin-
ical data of the inadequacy of the KW test since
patients with a genotypic score equal to 4 (Figure 2A)
had a larger median HIV-1 RNA reduction than
patients with a score equal to 3.

JT test
As previously, the first series of analyses considered
only mutation set I – associated with a poorer virolog-
ic response. The ‘adding’ procedure leads
to the selection of the combination of
M41L–M184VI+L74V, with a corresponding score taking the value –1, 0 or 1 and associated $P$ value of $2.11 \times 10^{-5}$. The ‘removing’ procedure selects
M41L+D67N–M184VI+L210W+K219Q/E to define
the genotypic score with a corresponding $P$ value of
$2.02 \times 10^{-7}$. HIV-1 RNA reduction for the two genotypic
scores provided from both original sets I and II using
the ‘removing’ procedure are displayed in Figure 2.
This figure provides an illustration with genuine clin-
ical data of the inadequacy of the KW test since
patients with a genotypic score equal to 4 (Figure 2A)
had a larger median HIV-1 RNA reduction than
patients with a score equal to 3.

<table>
<thead>
<tr>
<th>Position</th>
<th>Amino acid</th>
<th>$n$</th>
<th>Median decrease in HIV-1 RNA</th>
<th>$P$ value (Wilcoxon–Mann–Witney)</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>M</td>
<td>53</td>
<td>–0.85</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>49</td>
<td>–0.28</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>T</td>
<td>47</td>
<td>–0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y, F</td>
<td>55</td>
<td>–0.38</td>
<td>0.0001</td>
</tr>
<tr>
<td>210</td>
<td>L</td>
<td>74</td>
<td>–0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>28</td>
<td>–0.34</td>
<td>0.003</td>
</tr>
<tr>
<td>74</td>
<td>L</td>
<td>93</td>
<td>–0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>9</td>
<td>–0.06</td>
<td>0.005</td>
</tr>
<tr>
<td>67</td>
<td>D</td>
<td>67</td>
<td>–0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>35</td>
<td>–0.26</td>
<td>0.005</td>
</tr>
<tr>
<td>69</td>
<td>T</td>
<td>93</td>
<td>–0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>9</td>
<td>–0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>118</td>
<td>V</td>
<td>83</td>
<td>–0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>19</td>
<td>–0.28</td>
<td>0.17</td>
</tr>
<tr>
<td>219</td>
<td>K</td>
<td>77</td>
<td>–0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q, E</td>
<td>25</td>
<td>–0.44</td>
<td>0.20</td>
</tr>
<tr>
<td>70</td>
<td>K</td>
<td>75</td>
<td>–0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>27</td>
<td>–0.94</td>
<td>0.014</td>
</tr>
<tr>
<td>184</td>
<td>M</td>
<td>8</td>
<td>–0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>94</td>
<td>–0.61</td>
<td>0.042</td>
</tr>
</tbody>
</table>
score: M41L+D67N–M184V/I with a corresponding $P$ value of $7.08 \times 10^{-9}$. The ‘removing’ procedure selects the combination of M41L+T69D–K70R+L74V–M184V/I+T215Y/F+K219Q/E with a corresponding $P$ value of $4.52 \times 10^{-9}$. As expected, the magnitude of virological response is ordered according to the genotypic score value. HIV-1 RNA reduction for both genotypic scores obtained from both set I and set II using the ‘removing’ procedure are displayed in Figure 3. Only one patient had six mutations, which had a very small impact on the JT test (Figure 3A).

Table 2 summarizes the results of the different procedures for both original mutation sets I and II. One can remark that (i) whatever the original set of mutations and the procedure used, the JT test selects the combination providing the lower $P$ value (strongest association with virological response) compared with the KW test; (ii) given both original sets of mutations and test, the ‘removing’ procedure, starting from $K$ mutations, provides combinations including the larger number of mutations with smaller $P$ values than the ‘adding’ procedure except with set I using the JT test, where the combination of mutations is identical; and (iii) the use of mutation set II provides lower $P$ values than the use of set I.

Impact of unselected mutations
We investigated the impact of ‘unselected’ mutations to predict HIV-1 RNA reduction for genotypic scores using the JT test. Unselected mutations are the mutations that were selected by the univariate procedure (present in at least 5% of the samples with $P<0.20$) but not retained in the final scores, that is, mutations from Table 1 that are not retained in the final scores. Those mutations are added in a similar way to the way that genotypic scores are defined and their impact on virological response is investigated at each level of the corresponding score. Results are only displayed for original set II using the ‘adding’ procedure (Table 3) or ‘removing’ procedure (Table 4). For example, among the 11 patients with genotypic scores equal to $-2$, seven, three and one patients had none, one and two
### Table 2. Combination of mutations defining genotypic scores according to the mutations set, the non-parametric test and the procedure used

<table>
<thead>
<tr>
<th>Selection procedure</th>
<th>Non-parametric test</th>
<th>Set of mutations</th>
<th>Adding</th>
<th>Removing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KW</td>
<td>set I</td>
<td>M41L (P = 1.1 x 10^{-5})</td>
<td>M41L+V118I+L210W+T215Y/F+K219Q/E (P = 5.02 x 10^{-2})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>set II</td>
<td>M41L+L74V–M184V/I (P = 2.1 x 10^{-6})</td>
<td>M41L+D67N–M184V/I+L210W+K219Q/E (P = 2.02 x 10^{-7})</td>
</tr>
<tr>
<td></td>
<td>JT</td>
<td>set I</td>
<td>M41L+T69D+L74V+L210W+T215Y/F+K219Q/E (P = 1.17 x 10^{-7})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>set II</td>
<td>M41L+D67N–K70R+L74V–M184V/I (P = 7.08 x 10^{-9})</td>
<td>M41L+T69D–K70R+L74V–M184V/I+T215Y/F+K219Q/E (P = 4.52 x 10^{-9})</td>
</tr>
</tbody>
</table>

JT, Jonckheere–Tepstra; KW, Kruskal–Wallis.

### Table 3. Impact of unselected mutations in the final genotypic score using mutation set II, the JT test for trend and the ‘adding’ procedure

<table>
<thead>
<tr>
<th>Unselected mutations (T69D+V118I+L210W+T215Y/F+K219Q/E)</th>
<th>Genotypic score (M41L+D67N–K70R+L74V–M184V/I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.23 (7)</td>
</tr>
<tr>
<td>1</td>
<td>0.99 (3)</td>
</tr>
<tr>
<td>2</td>
<td>0.38 (1)</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
</tr>
</tbody>
</table>

P value (JT test): 0.04 0.37 0.21 0.43 0.44

JT, Jonckheere–Tepstra.

### Table 4. Impact of unselected mutations in the final genotypic score using mutation set II, the JT test for trend and the ‘removing’ procedure

<table>
<thead>
<tr>
<th>Unselected mutations (D67N+V118I+L210W)</th>
<th>Genotypic score (M41L+T69D–K70R+L74V–M184V/I+T215Y/F+K219Q/E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.23 (7)</td>
</tr>
<tr>
<td>1</td>
<td>1.39 (1)</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
</tr>
</tbody>
</table>

P value (JT test): 0.26 0.39 0.33 0.11 0.15 0.42

JT, Jonckheere–Tepstra.
mutations (mutations among T69D, V118I, L210W, T215Y/F and K219Q/E), respectively, with a median HIV-1 RNA reduction of 1.23, 0.99 and 0.38, respectively (Table 3). The $P$ value of 0.04 indicates that significant information on HIV-1 RNA decrease is contained in the sum of the unselected mutations. All other $P$ values, however, are not statistically significant. Unselected mutations are also not associated with the virological response for each level of the genotypic score obtained using the ‘removing’ procedure (Table 4).

### Discussion

We described and compared the use of two non-parametric tests and procedures for the process of selecting mutations to build clinically relevant rules of genotype. The use of genuine clinical data indicated the superiority of the JT test for selecting the appropriate combination of mutations most strongly associated with the HIV-1 RNA reduction. This finding was expected since the JT test is specifically designed to compare ordered alternatives, such as groups of patients with distinct numbers of mutations, taking into account that mutations associated with a better response have a negative weight. Results also indicate that the procedure starting with the $K$ mutations retained and removing mutations one by one, provides genotypical scores more strongly associated with virological response than starting with a single mutation and adding mutations one by one. The former technique tends to select a higher number of mutations than the latter technique. We also suggest a simple way of investigating whether unselected mutations provide information on virological response conditionally on levels of the genotypical score. Unselected mutations in the removing procedure were not significantly associated with virological response in the Jaguar trial.

Other methods can be used to build guidelines for genotype rules or interpretation. Methods such as discriminant analysis, recursive partitioning or cluster analysis have been suggested to correlate drug-susceptibility phenotype and genotype, although they can be used to correlate genotype and virological response [10–14,24]. These methods often provide similar results. Simple linear regression models, however, can also be envisaged in this setting. The pre-selection of mutations eligible for genotypic score can be carried out in a similar way to that described in our article with a series of univariate linear regression. The linear model can then be used in two slightly different ways. The first possibility is to consider a multivariate model that estimates a parameter for each mutation. Both forward selection and backward elimination techniques can be used to select the final multivariate model that will provide a set of mutations, each one being independently associated with the virological response. This approach will not provide a genotypical score as described in this work. Each parameter of the final model can be interpreted as the weight of the corresponding mutation, conditionally on the set of other mutations involved in the model, to predict HIV-1 RNA reduction. The problem with this approach is the high degree of co-linearity between these mutations so that reliably estimating the parameters is difficult.

The second possibility is to consider, as in our procedure, a variable representing a sum of mutations and to estimate the associated parameter to this variable in a linear regression. Both ‘adding’ and ‘removing’ procedures can be used, as described in the Methods section. The linear constraint of the model assumes here that the effect on HIV-1 RNA reduction is exactly the same between patients having none or one mutation than, for example, in patients having two or three mutations. Such an approach is close to our approach but more restrictive since the JT test only implies that, on average, the HIV-1 RNA reduction will be lower for groups of higher numbers of mutations without any strict linear correspondence.

The approach used in this work considers that mutations defining the genotypic score have the same weight in predicting the HIV-1 RNA reduction. It is clear, however, than some mutations have a distinct weight in drug resistance. For example, in the ANRS algorithm, viruses are considered as resistant to zidovudine (ZDV) either in presence of T215Y/F or with at least three mutations amongst M41L, D67N, K70R, L210W and K219Q/E. In this case, T215YF has a weight three times higher than each of the five other mutations. The weight of each mutation can be estimated in a regression model. Databases are required, however, including large numbers of patients harbouring a single mutation to estimate the weight of each mutation reliably. This difficulty has encouraged a wide collaboration between different people and groups such as the HIV Resistance Response Database Initiative (http://www.hivrdi.org) and the Forum For Collaborative HIV Research (http://www.hivforum.org).

It is well known from statisticians that in multivariate modelling, selection techniques lead to different results whatever the regression model used. Therefore, it is not surprising that the two procedures used in a non-parametric setting provide different genotypic scores. In addition, the two procedures have two slightly different objectives. The objective of the ‘adding’ procedure is to find the minimum number of mutations allowing the split of patient groups explaining the maximum of the HIV-1 RNA reduction. The objective of the ‘removing’ procedure is, from a sum including all pre-selected mutations, to keep a
mutation that significantly contributes to split groups of patients. The latter technique will automatically provide a genotypic score with larger number of mutations than the former technique. An extreme example is given in Table 2, where using the KW test and mutation set I, the first procedure provides a genotypic 'score' defined with only one mutation (M41L).

The approach of adding mutations without any specific weight for each mutation can be considered as too simplistic to describe the complexity of the mechanism for drug resistance. We must consider that mutations have distinct weights with respect to virological response, and also that interactions between mutations play a role in the drug resistance pattern as, for example, in the different pathways described for ZDV resistance [25,26]. A specificity of the Jaguar study is the low number of patients (11%) in the ddI arm having an HIV-1 RNA reduction censored by the limit of quantification (LOQ) of the assay (HIV-1 RNA <50 copies/ml). The reason is the short duration of the trial since the primary endpoint was the HIV-1 RNA reduction at week 4 and probably because a single active drug (ddI) was initiated. In other trials, however, a high percent of HIV-1 RNA is censored by the LOQ and survival methods should be used to estimate HIV-1 RNA reduction [19–21]. In this context, the JT non-parametric test as well as the KW, as described in this manuscript, cannot be used. The JT test, however, has been extended to the censored case generalizing the weighted log-rank approach [27]. A test based on the logit-rank procedure has also been suggested, which generalizes the JT test in the case of censored observations [28].

Tests and procedures should be applied to other data to confirm or inform findings introduced in the present work. In particular, further work is needed to fully compare the two selection procedures. Each procedure has advantages and shortcomings that are related to objectives of the analysis. In the Jaguar study, P values are extremely low, which is due to the add-on design of the study, while lower P values are obtained with the JT test which is more powerful than the KW test. The criterion used to end the two selection procedures was to obtain a larger P value than the one provided by the previous combination of mutations.

References


HIV treatment response from genotype may depend on diversity as well as size of data sets. 11th Conference on Retroviruses and Opportunistic Infections, 8–11 February 2004, San Francisco, CA, USA. Abstract 697.


