Validation of the revised 10-item HIV Treatment Satisfaction Questionnaire status version (HIVTSQs) and new change version (HIVTSQc)

Publication: Woodcock A and Bradley C (2006) Validation of the revised 10-item HIV Treatment Satisfaction Questionnaire status version (HIVTSQs) and new change version (HIVTSQc). *Value in Health* **9** (5) 320-333.
http://www.ncbi.nlm.nih.gov/pubmed/16961550

Authors: Alison Woodcock, Ph.D, C.Psychol, AFBPsS (corresponding author)

Clare Bradley, Ph.D, C.Psychol, FBPsS, FRSM.

Affiliation (both authors): Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham, Surrey, England, TW20 0EX.

Address for correspondence: Dr Alison Woodcock, Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham, Surrey, TW20 0EX.. England, UK.

Fax: +44-1784 434347 **email:** a.woodcock@rhul.ac.uk

Funding: This work was funded by a research grant from GlaxoSmithKline.

Key words: HIV, questionnaire, treatment_satisfaction, ethnicity.

Running title: HIV Treatment Satisfaction Questionnaire (status and change)

Abstract

Objectives: Previous validation of the HIV Treatment Satisfaction Questionnaire status version (HIVTSQs) found that 9/10 items performed well, but the *demands* item needed revision. This study investigated the psychometric properties of the revised HIVTSQs and new change version (HIVTSQc).

Methods: English-speaking Americans completed the HIVTSQs at baseline and Week 48 of a clinical trial of HIV treatments, and the HIVTSQc at Week 48. Demographic and viral load information was collected. Psychometric validation used item frequency distributions, Confirmatory Factor Analysis (CFA), item-total correlations, Cronbach's alpha, Spearman's rank correlation, Kruskal-Wallis and Mann-Whitney tests.

Results: At baseline, 126/152 patients completed the HIVTSQs fully (100/106 at Week 48). The negatively skewed distribution of the revised *demands* item resembled that of the other nine, with comparable missing data. CFA (baseline and Week 48) supported the *general satisfaction/clinical* subscale (alpha 0.83; 0.85), *lifestyle/ease* subscale including *demands* (alpha 0.82; 0.85), and ten-item *treatment satisfaction* scale (alpha 0.89; 0.91). Subscale and scale scores differed significantly between ethnic groups. Viral load was not significantly related to subscale or scale scores. At Week 48, 97/106 patients completed the HIVTSQc fully. All items had negatively skewed distributions. CFA supported two subscales (*general satisfaction/clinical change* alpha 0.85; *lifestyle/ease change* alpha 0.88) and ten-item *treatment satisfaction change* scale (alpha 0.92). Viral load change between baseline and Week 48 correlated significantly with patients' perceptions of change in HIV control (*control(c)* item), but not with scale or subscale scores.

Conclusions: The excellent psychometric properties of the HIVTSQs and c support their use in clinical trials. 250

Introduction

The HIV Treatment Satisfaction Questionnaire (HIVTSQ) was developed to evaluate treatments for Human Immunodeficiency Virus (HIV) [1]. It is one of several condition-specific measures using the format of the Diabetes Treatment Satisfaction Questionnaire (DTSQ) [2],[3]. Others are for end stage renal disease [4], genital herpes [5], diabetic retinopathy [6], and hypothyroidism [7], as well as the DTSQ-Teen for teenagers with diabetes and DTSQ-Parent for parents of children with diabetes.

In the original validation study, 150 American and Canadian patients completed an earlier version of the ten-item HIVTSQ status version (HIVTSQs). Psychometric analyses [1] revealed two subscales (the five-item *general satisfaction*/*clinical* subscale (range 0-30) and the four-item *lifestyle/ease* subscale (range 0-24)) as well as the total *treatment satisfaction* scale (range 0-54), computed from nine of the ten items. Each scale and subscale had good internal consistency reliability. The validation showed that the fourth item, *How demanding is your present form of treatment (in terms of time, effort, thought etc.)?* was problematic. It was the only item for which the least favourable response *(very demanding)* was to the left of the scale, scoring 6 and the most favourable response *(very undemanding)* to the right, scoring 0. For the remaining nine items, the most favourable response was to the left. Distributions of nine items were negatively skewed, most patients indicating high levels of satisfaction. Scores for the *demanding* item, however, formed a more rectangular distribution, indicating that some patients may have circled responses towards the left, believing these to be more favourable, as for the other nine items. Statistical analyses showed that the *demanding* item did not fit into the structure of the measure. Recommendations for computing scores therefore excluded this item. The report recommended that the problematic item wording should be changed to: *How satisfied are you with the demands made by your current treatment? (*response options *very satisfied* (6) to the left, to *very dissatisfied* (0) to the right). The earlier version of the HIVTSQs had already been used in two further trials and translated into several languages without involvement of its authors. For those language versions with acceptable translations, psychometric analyses strongly supported the nine-item *treatment satisfaction* scale, which proved sensitive to differences between treatment groups in three trials [8].

A ‘change’ version of the DTSQ (the DTSQc) has been found valuable in overcoming ceiling effects that often arise with status measures of satisfaction, the DTSQs (status) version being no exception. Although the DTSQs has proved sensitive to change in many trials (reviews [2],[9]), the DTSQc enhances that sensitivity by allowing those who were satisfied at baseline to express even greater satisfaction at follow-up [10],[11]. An HIVTSQ change version should prove similarly useful in identifying improvements, particularly where ceiling effects are apparent.

Figure 1

Linguistically validated versions of the HIVTSQs and a change version, the HIVTSQc (Figure 1) were therefore prepared in several languages, each with the revised *demanding* item having the most favourable response to the left (now labeled *demands,* item 4 in each measure). In the present work we investigate the structure of the revised HIVTSQs and the new HIVTSQc (English version completed in USA) and the internal consistency reliability of any scales/subscales, to determine the scoring for each measure, including procedures for dealing with missing values and the relationship between questionnaire scores, demographic variables and viral load.

Methods

Procedure

At baseline (Day 1) of a clinical trial of an investigational product for treatment of HIV, just before randomisation to one of three treatments, the HIVTSQs (English version) was administered to 152 English-speaking American, HIV-positive individuals already receiving antiretroviral therapy. This was a different sample from the 150 patients in the previous validation study [1]. At screening, which preceded baseline by some weeks, patients eligible to participate in the present trial were ‘experiencing failure of their antiretroviral therapy regimen’ (plasma HIV-1 RNA ≥1,000 copies/mL). All patients who completed at least one HIVTSQ measure during the trial were included, whether or not they completed the trial. The HIVTSQc (English) was administered to the 106 patients remaining in the trial at Week 48 on randomised treatment. Patients completed the measures themselves, before seeing the clinician and before being informed of their viral load. Demographic data, including age, gender and ethnicity were recorded, as was viral load at each time point.

Analysis

Demographic and disease characteristics were summarised. The distribution of item scores and frequency of missing data were examined. Week 48 HIVTSQc item scores were correlated against the difference between HIVTSQs item scores at baseline and Week 48 (computed as Week 48 - baseline). Analyses followed broadly the same sequence for both HIVTSQs and c. Confirmatory Factor Analysis (CFA) was conducted on baseline and Week 48 HIVTSQs item scores (without and then with the *demands* item),and on all ten HIVTSQc scores, using the maximum likelihood estimation algorithm of Amos 4.01, to determine whether the *general satisfaction/clinical* and *lifestyle/ease* subscales, then the *treatment satisfaction* scale were supported. A normed fit index (NFI) ≥0.9 was considered to indicate an acceptable fit [12, pages 407-8] and standardised regression weights >0.4 for contributing items were considered acceptable. In SPSS for Windows 12.0, corrected item-total correlations (each item correlated with the sum of the remaining items) and Cronbach's alpha were used to determine the internal consistency reliability of scales/subscales identified in the CFAs. Alpha >0.7 is usually considered acceptable, though for some purposes, alpha >0.8 is desirable. Item-total correlations and alpha with each item removed were examined to identify any items not contributing well to the scale, indicated by a relatively poor item-total correlation and an increased alpha if that item was removed.

The method for computing missing values for any scale or subscale involved removing first the item whose removal caused the greatest fall in the scale alpha. This is the 'strongest' item and internal consistency will be damaged most if this item is missed. Cronbach's alpha was then calculated without that item. If the alpha remained above 0.7, the item whose removal *now* caused the greatest fall in the scale alpha was removed and Cronbach's alpha calculated again without both the strongest and second item. The process was repeated cyclically until the scale alpha fell below 0.7. The number of items that could be missed, but the scale score still computed using completed items, was the number of items remaining in the Cronbach's alpha analysis before alpha fell below 0.7. The number of items that could be missed whilst retaining an alpha >0.8 was also ascertained. In order to retain a range of content when judging satisfaction, a maximum of half the items may be missed (e.g. five from a ten-item scale), regardless of the alpha achieved.

Scale and subscale scores were then computed and associations with demographic characteristics (age at recruitment, sex and ethnicity) and viral load at that visit were examined. Non-parametric

statistics were used to determine whether there were significant demographic differences: Mann-Whitney U tests compared the two age categories (median split into younger <40 and older ≥40 years) and the two sexes; Kruskal-Wallis multiple comparison test, backed by Mann-Whitney U tests, compared three ethnic groups (Hispanic, African-American, White). In each case, a probability (p) of <0.05 (two-tailed) indicated a significant difference between subgroups. Subscale and scale scores and *control* item scores were correlated against viral load (Spearman's rank correlation). In the case of the HIVTSQs, this was viral load at baseline and again at Week 48; in the case of HIVTSQc, it was change in viral load between baseline and Week 48, computed as (Week 48 – baseline) so that a more negative score represented a greater decrease in viral load during the trial (improved HIV control).

Results

HIVTSQs Validation

Demographic and disease characteristics

Table 1

Table 1 describes the HIVTSQs baseline sample. The majority were men and the average age was around 40 years. Just over half were White and 37.5% were African-American. Viral load varied considerably, shown by the minimum and maximum values and the fact that the standard deviation exceeded the mean. The viral load distribution was positively skewed, 51.1% with viral load <11,000 and only 7.1% with viral load >200,000. HIV-1 RNA(log_{10}) is also shown. Although all were immunologically compromised when screened for trial participation (HIV-1 RNA ≥1,000 copies/mL), twelve had viral loads <1,000 and three had viral loads <400 HIV-1 RNA copies/mL at baseline. One scored the lowest possible on the assay (<50, coded as 49).

Distribution of HIVTSQs item scores

Scores for every HIVTSQs item, including *demands*, had a negatively skewed distribution, with a greater frequency of responses towards the higher end of the 0-6 scale, indicating generally high satisfaction. Baseline satisfaction scores did, however, include some towards the centre and lower end of the scale, as expected for people experiencing failure of antiretroviral therapy. *Control* item

scores most approximated a normal distribution (mode = 4). For the remaining items, the modal score was higher (mode = 5 for items 4, 5, 6 and 8; mode = 6 for items 1, 3, 7, 9 and 10).

HIVTSQs missing values

If any item is missed more than the others, the wording may be problematic or the content irrelevant to the patient population. 126 people (82.9%) completed the HIVTSQs in full at baseline. Although 10 missed the HIVTSQs entirely, it is not known whether they were given it to complete. A further 16 missed some items, but not others (between 1 and 8 items missed) and so had clearly been given the HIVTSQs to complete. The items they missed were examined to determine whether the revised *demands* item was missed any more frequently than other items. The number to miss each item ranged between 2 *(lifestyle)* and 9 *(understanding)*. Only 4 missed *demands.* Demographic characteristics of those who missed items were examined. Only 77.2% of African-American people completed the HIVTSQs fully, compared with 85.7% Hispanic and 86.3% White people. Those missing all ten items were similar demographically to the entire sample: 9 men and 1 woman; 7 White, 2 African-American and 1 Hispanic; mean age 44.5 (sd 6.3). Of the16 who completed some items but not others, however, 11 were African-American and only 4 were White. All those missing 6+ items were African-American. Considering the 142 who attempted the HIVTSQs, 69/73 White patients, 12/13 Hispanic and the 'Other' man completed it fully, but only 44/55 African-American patients. A χ^2 test (three ethnic groups (excluding the 'Other' man) and two completion categories (full versus partial completion)) revealed a significant ethnic difference in completion ($\chi^2 = 6.76$; n = 141; df = 2; p = 0.03).

HIVTSQs Structure

Baseline structure

Table 2

The hypothesised structure found in the original validation [1], with items 1,2,3,9 and 10 loading on one factor (*general satisfaction/clinical)*, and items 5-8 loading on the other (*lifestyle/ease*), was tested, allowing for correlation between the two latent variables, using the maximum likelihood estimation algorithm in Amos. The fit of the model was good: X^2 for the solution was 56.81 (df = 26; $p<0.001$) and $X^2/df = 2.19$ and normed fit index (NFI) = 0.98, which exceeds the required value of 0.9 for goodness of fit. The two latent variables correlated 0.69. When the *demands* item was added

(hypothesised as fitting into the *lifestyle/ease* subscale), X^2 was 116.67 (df = 34; p<0.001), X^2/df = 3.43 and NFI = 0.97. The two latent variables correlated 0.76. In both the nine-item and ten-item analysis, nine items had standardised regression weights >0.4, but the weight for *understanding* was slightly lower. Thus, the fit to the model with two five-item subscales (*general satisfaction/clinical* and *lifestyle/ease)* was acceptable, but *understanding* did not fit as well as the other items. When the goodness of fit of a single nine-item scale model (without *demands*) was tested, X^2 was 116.73 (df = 27; p<0.001), X^2/df = 4.32 and NFI = 0.96. Adding in the *demands* item, X^2 was 167.37 (df = 35; p<0.001), X^2/df = 4.78 and NFI = 0.95. In both cases, the standardized regression weight for *understanding* fell just below 0.4, but the weight for *demands* was strong (0.85) (see Table 2 summarising the 10-item CFAs)

Week 48 structure

In order to check this structure amongst the same participants but with experience of different trial treatments, CFA was performed on data collected in Week 48 for all patients completing the trial on randomised treatment (even though some changed their background therapy for reasons of safety or intolerance). Because the three treatment groups had different treatment experiences, the following check was conducted to determine whether their HIVTSQs scores could be combined for psychometric analysis: (i) The raw scores for the ten items were subjected to Principal Components analysis (PCA), with a forced single-factor solution. (ii) Within each of the three treatment groups separately, normalised scores (z-scores) were computed for each questionnaire item. Combining the dataset again, these z-scores were subjected to PCA, with a forced single-factor solution. (iii) The ten factor loadings for (i) were regressed against those from (ii) (n = 10). If the regression coefficient is close to 1 and the constant close to zero then the three subsets of data can be combined. The correlation between the two sets of loadings was very high (R = 0.994), the constant (-0.050) was not significantly different from zero [t(8) = 1.57; p>0.05] and the slope (1.061) was not significantly different from 1 [t (8) = 1.419; p>0.05]. This indicated that even though the absolute values in each treatment group might differ, the inter-relationships between items are similar and the three groups can be combined for further analysis. All ten HIVTSQs items (Week 48) were entered into a CFA, to determine goodness of fit of the two-subscale model. X^2 was 83.23 (df = 34; p<0.001), X^2/df = 2.45 and NFI = 0.98, with the two latent variables correlating 0.83. All ten items had standardized

regression weights >0.4, though *understanding* was the weakest at 0.41. Testing the goodness of fit of a ten-item single scale model (Week 48), X^2 was 137.95 (df 35; p<0.001), X^2/df = 3.94 and NFI = 0.97. Nine items had standardized regression weights >0.4, though *understanding* was exactly 0.4 (Table 2).

Internal consistency reliability of two subscales and single scale

Table 3

Internal consistency reliability analysis was conducted first on baseline data.Table 3 provides the corrected item-total correlation for each item and alpha with each item removed (n = 129 who completed all five *general satisfaction/clinical* subscale items, n = 128 for the five *lifestyle/ease* items and n = 126 for the ten *treatment satisfaction* scale items). The scale alpha for *general satisfaction/clinical* was 0.83, for *lifestyle/ease*, α was 0.82 and for *treatment satisfaction*, α was 0.89. In each case, all item-total correlations were >0.4. As expected from the CFA, the weakest item was *understanding* in both the *general satisfaction/clinical* subscale and *treatment satisfaction* scale. Alpha >0.8 confirmed very good internal consistency reliability at Week 48 (*general satisfaction/clinical* α = 0.85; lifestyle/ease α = 0.85 and *treatment satisfaction* α = 0.91). We conclude from these analyses that the *general satisfaction/clinical* subscale can be computed from items 1,2,3, 9 and 10 and the *lifestyle/ease* subscale from items 4-8 (possible range 0-30 for each subscale) and *treatment satisfaction* scale score can be computed as the sum of all ten HIVTSQs items (possible range 0-60), but that *understanding* is a relatively weak item..

Dealing with missing HIVTSQs values

Items were removed cumulatively and alpha calculated again for each subscale and scale at baseline. Their removal was not in the exact order expected from the original alpha-if-item-deleted values. For example, item 1 was removed from the *treatment satisfaction* scale before item 5. This is because the number of participants contributing to the analysis increased and the content of the scale changed with cumulative removal of items. Removing the strongest item (*current treatment*) from the *general satisfaction/clinical* subscale resulted in a fall to α = 0.767, but removal of the second strongest (*continue*) would result in a fall to 0.681, so only one *general satisfaction/clinical* item can be missed and substitute for with the mean of the remaining four items before the *general satisfaction/clinical* subscale is computed as the sum of the five items. Removing the strongest item (*convenience*) from the *lifestyle/ease* subscale resulted in a fall to α = 0.742, but removal of the

second strongest (*lifestyle*) would result in a fall to 0.606, so only one *lifestyle/ease* item can be missed and substituted by the mean of the remaining four items before the *lifestyle/ease* subscale is computed as the sum of the five items. Each subscale cannot be computed for those missing more than one contributing item. If greater internal consistency reliability of >0.8 is needed, then no missing values can be tolerated. Removal of the strongest item from the total *treatment satisfaction* scale, (*demands*) reduced α to 0.868. Removal also of *current treatment* reduced α to 0.847, then additional removal of *convenience* reduced α to 0.819. Thus, three items may be missed before alpha falls below 0.8. Further removal of *recommend to others* reduced α to 0.778 and removal of *control* reduced α further to 0.723. Removal of *understanding* would reduce the scale alpha to 0.705, which just exceeds the criterion level of α>0.7. However, in order to retain the range of content of the measure when computing missing values, a maximum of five items (half the scale) may be missed and each computed as the mean of completed item scores. The *treatment satisfaction* scale score can now be computed as the sum of the ten item scores, including substituted means as necessary for up to five items (for α>0.7) or three items (for α>0.8). If six or more items are missed, the scale score should not be computed.

Computing two subscale and scale scores

Using the method of computing subscales scores, in which only one missing value is replaced by the mean, 136 *general satisfaction/clinical* subscale scores were computed at baseline (an additional seven, compared with the 129 completing all five items). Mean *general satisfaction/clinical* was 20.6 (sd 6.7), median and mode 22, ranging 2 to 30 and skewness -0.53 (se 0.21). Using the same method for the *lifestyle/ease* subscale, 137 scores could be computed (compared with 128 completing all five items). Mean lifestyle/ease was 22.9 (sd 5.6) median and mode 24, ranging 1-30. Thus, both subscales provide a range of scores, with a negative skew, the median being greater than the mean. Using the method of computing missing values in which up to five missing values can be replaced by the mean, baseline *treatment satisfaction* scores were computed for 139 patients, an additional 13 (eight African-American, four White and one Hispanic) compared with the 126 who completed the HIVTSQs fully. Mean *treatment satisfaction* was 43.5 (sd 11.3); median 45 (ranging 5 to 60). The mode was 50 (10 below the maximum possible) and skewness -0.66 (se 0.21).

Examination of subscale and scale scores in relation to demographic variables and viral load

Table 4

There was no significant difference between the two age categories or between sexes in *general satisfaction/clinical*, *lifestyle/ease* or *treatment satisfaction*, but there was a difference (Kruskal-Wallis) between ethnic groups for all three scores (Table 4). At baseline, African-American people reported significantly higher *general satisfaction/clinical* than did White people (Mann-Whitney p = 0.034), higher *lifestyle/ease* (p = 0.011) and higher *treatment satisfaction* (p = 0.007). The direction of the difference is all the more striking when baseline viral load is examined. Although there was no significant difference between viral loads of the three ethnic groups (F = 1.82; df 2; p = 0.17), the mean viral load for White patients was 38,540.6 HIV-1 RNA copies/mL (sd 91,958.1), whereas for African-American patients, it was higher at 77,868.1 (sd 144,729.2), the Hispanic patients having an intermediate mean 62,215.4 (sd 97768.7). African-Americans thus had poorer HIV control, yet greater satisfaction.

Baseline viral load was correlated against *control* item scores (r = -0.14; n = 138; p = 0.09), against *general satisfaction/clinical* (r = -0.12; n = 135; p = 0.16), *lifestyle/ease* (r = -0.003; n = 136; p = 0.98) and *treatment satisfaction* (r = -0.09; n = 138; p = 0.60). None were significant. At Week 48, however, the correlation with viral load was significant for the *control* item (r = -0.44; n = 102; p<0.001) and for *general satisfaction/clinical* (r = -0.24; n = 101; p = 0.18) whilst remaining non-significant for *lifestyle/ease* (r = 0.001; n = 102; p = 0.99) and *treatment satisfaction* (r = -0.10; n = 102; p = 0.30). Investigation of baseline correlations with viral load for each ethnic group (Hispanic, African-American and White) revealed a significant correlation between viral load and the *control* item for White patients (r = -0.23; n = 73; p = 0.046), but this was weaker for African-American people (r = -0.17; n = 51; p = 0.22) and negligible for Hispanic patients (0.08; n = 13; p = 0.80). At Week 48, the correlation was more strongly significant for White patients (r = -0.42; n = 56; p = 0.001), significant for African-American patients (r = -0.48; n = 36; p = 0.003) and followed the same pattern, albeit less strongly, for the much smaller sample of Hispanic patients (r = -0.13; n = 9; p = 0.74). The negative correlation between viral load and *general satisfaction/clinical* scores was stronger at Week 48 for African-Americans, (r = -0.34; n = 35; p = 0.048) and for White patients (r = -0.17; n=56; p=0.20) than it had been at baseline (African-American (r = -0.21; n = 49; p= 0.14); White (r = -0.10; n = 72; p = 0.38), indicating a stronger relationship between patient perceptions of the clinical effectiveness of their treatment and test results as the trial progessed.

HIVTSQc Validation

The analyses below used HIVTSQc data collected at Week 48 from patients completing the trial on their randomised treatment.

Demographic and disease characteristics

By Week 48, 45 patients had withdrawn, leaving 107. However, a Hispanic man completed the English HIVTSQs at baseline, but US Spanish versions of HIVTSQs and c thereafter, and he is therefore excluded from Week 48 HIVTSQc analysis, leaving n=106 (Table 1). When recruited, the subset of the sample providing HIVTSQc data were about a year older than were the larger HIVTSQs baseline sample, indicating that younger people may have withdrawn early from the trial. Ethnic groups were in similar proportions to the HIVTSQs baseline sample. At Week 48, average viral load was lower than in the baseline analysis: 57 patients (54.3% of 105 with viral load results) had the minimum level detected by the assay (coded as 49 HIV-1 RNA copies/mL), compared with only one person at baseline; furthermore, 83/105 (79%) had a viral load <1,000, compared with 12 at baseline.

Distribution of HIVTSQc item scores

Distribution of scores for the ten HIVTSQc items was examined within each of the three treatment groups. The distribution for all ten items was negatively skewed, with a modal response of +3 for almost all items, indicating the maximum possible degree of improvement (e.g. *much more satisfied now*). The only exceptions were the mode of +2 for *side effects(c)* in treatment groups 1 and 3, mode of +2 for *understanding(c)* and joint modal score of 2 and 3 for *lifestyle(c)* in treatment group 1. The minimum score for *control(c)* was zero for all three treatment groups, indicating that all patients thought their HIV control was either as good as before the trial, or better. However, for each of the other nine items, a few patients indicated a negative change. For *demands(c), flexibility(c), lifestyle(c)* and *recommend to others(c)*, there were negative changes for at least one person in each treatment group, indicating a belief that the trial treatment was less satisfactory than their previous treatment. Spearman's correlation between Week 48 change item scores and the difference between status scores obtained at baseline and Week 48 were all positive (ranging between r = +0.12 and +0.39).

Five correlations were significant ($p<0.05$), indicating that change judgments related to differences in absolute judgments of satisfaction, though not very strongly.

HIVTSQc missing values

Of the 106 who attended at Week 48,, only nine missed any HIVTSQc items, of whom two missed the entire measure. It is not certain that those two were given the HIVTSQc to complete. However, the seven people missing 1-9 items clearly had received the measure and each item was missed by between 2 and 4 of these people. Thus, no item was missed any more frequently than another. Whilst 56 (98.2%) White patients and the 'Other' ethnicity patient completed the measure fully, only 8 (88.9%) Hispanic and 32 (82.1%) African-American patients did so.

HIVTSQc Structure

Because the three treatment groups had different treatment experiences, the raw-score z-score check was conducted to determine whether their HIVTSQc scores could be combined for psychometric analysis, as described above for the Week 48 HIVTSQs data. The correlation between the two sets of loadings was very high ($R = 0.995$), the constant (-0.041) was not significantly different from zero [$t(8) = 1.530$; $p>0.05$] and the slope (0.950) was not significantly different from 1 [$t(8) = -1.429$; $p>0.05$]. This indicated that even though the absolute values in each group might differ, the inter-relationships between the items are similar and the three treatment groups can be combined for further analysis.

Table 5

CFA first sought to confirm the two-subscale model, by including all ten item scores in the analysis, with intercorrelation between the two latent variables. Each HIVTSQc change item was allocated to the same subscale as for the status version. X^2 was 193.60 ($df = 34$; $p<0.001$), $X^2/df = 5.69$ and NFI = 0.92 and the two latent variables correlated 0.85. For a single ten-item *treatment satisfaction change* scale (all items loading onto one latent variable), X^2 was 257.03 ($df = 35$; $p<0.001$) $X^2/df = 7.34$ and NFI = 0.90, which is just acceptable. For both the two-subscale and single scale model, all ten standardised regression weights were >0.4 (Table 5).

Internal consistency reliability of the treatment satisfaction change scale

Table 6

The internal consistency reliability of each of the the five-item subscales (*general satisfaction/clinical change* and *lifestyle/ease change*) and the ten-item *treatment satisfaction change* scale was tested,

using corrected item-total correlations and Cronbach's alpha statistics. Table 6 provides the item-total correlation for each item and alpha if each item is removed (n = 98 for each of the two subscales and n = 97 for the *treatment satisfaction change* scale, having completed all HIVTSQc items within the relevant scale/subscale at Week 48). The scale alpha for *general satisfaction/clinical change* was 0.85, the strongest item being *continue(c);* for *lifestyle/ease change*, alpha was 0.88, the strongest item being *lifestyle(c)* and for *treatment satisfaction change*, alpha was 0.92, the strongest item again being *continue(c)*. The *demands(c)* item fitted well into both the *lifestyle/ease change* subscale and the total *treatment satisfaction change* scale. The items least integral to the *treatment satisfaction change* scale were *control(c)* and *understanding(c)*. Alpha increased when *control(c)* was removed from the *general satisfaction/clinical change* subscale and when *understanding(c)*_was removed from the *lifestyle/ease change* subscale. The *general satisfaction/clinical change* subscale can be computed as the sum of change items 1,2,3,9 and 10, the *lifestyle/ease change* subscale computed as the sum of change items 4-8 (possible range -15 to +15) and the *treatment satisfaction change* score computed as the sum of all ten HIVTSQc items, (possible range +30 to -30). The higher the positive score, the greater is the improvement in satisfaction since the start of the trial. The more negative the score, the greater is the deterioration in satisfaction.

Dealing with missing HIVTSQc values

In the *general satisfaction/clinical change* subscale, removal of the strongest item (*continue(c))* caused alpha to fall to 0.768, but removal also of *current treatment(c)* would cause too great a fall to 0.548. In the *lifestyle/ease change* subscale, removal of the strongest item, *lifestyle(c)*, caused alpha to fall to 0.824, but removal then of item 6 (*flexibility(c)*) caused a fall to 0.696, (just below the 0.7 desired minimum). Therefore, only one item may be missed from each of the subscales and substituted by the mean of the four remaining items before summing all five to produce the relevant subscale score. Subscale scores cannot be computed for those missing more than one item from the subscale concerned. To retain an alpha >0.8, no missing items can be tolerated in either subscale. Removal of the strongest item from the *treatment satisfaction change* scale,, *lifestyle(c)*, reduced alpha to 0.897. The additional removal of *continue(c)* reduced it to 0.862, removal also of *flexibility(c)* reduced it to 0.829 and removal also of *current treatment(c)* reduced it to 0.801 (still above the higher criterion of 0.8). Removal of *demands(c)* reduced it considerably to 0.718. Removal of

understanding(c) would reduce it to 0.707, still above the criterion level of 0.7. However, in order to retain the range of content, only five items may be missed and each can be substituted by the mean of completed items (maximum of four items to retain alpha >0.8). The *treatment satisfaction change* score can then be computed as the sum of the ten item scores, using substituted means where necessary. If six or more items are missed, the scale score cannot be computed.

Computing subscale and scale change scores

Week 48 *general satisfaction/clinical change* and *lifestyle/ease change* scores were computed, using the method of substituting up to one missing value. For *general satisfaction/clinical change*, 99 scores were computed (one more than the 98 completing all five items). Mean *general satisfaction/clinical change* was 11.5 (sd 4.3), median 13, ranging from -5.0 to +15.0; mode +15 (the maximum possible positive change) and skewness -2.26 (se 0.24). For *lifestyle/ease change,* 100 scores were computed (two more than the 98 completing all five items). Mean *lifestyle/ease change* was 10.8 (sd 4.8), median 12, ranging from -6.0 to +15.0; mode +15 (the greatest possible positive change) and skewness -1.56 (se 0.24). *Treatment satisfaction change* scores, using the method of substituting up to a maximum of five missing values, could be computed for 100 patients (the 97 who completed the HIVTSQc in full, and the additional three who missed either one or two items. The mean *treatment satisfaction change* score was 22.6 (sd 8.6); median 26.0 (ranging -8.3 to +30). The distribution was negatively skewed (skewness -1.88 (se 0.2)) with a mode of +30 (the maximum possible positive change). Thus, both subscale and scale scores were skewed considerably.

Examination of subscale and scale change scores in relation to demographic variables and viral load

Table 7

There were no significant age, sex or ethnic differences in HIVTSQc *general satisfaction/clinical change, lifestyle/ease change* or *treatment satisfaction change* scores (Table 7). The correlation between change in viral load from baseline to Week 48 and *control(c)* item scores was small but significant (r = -0.23; n = 93; p = 0.03). The correlations between change in viral load from baseline to Week 48 and the Week 48 HIVTSQc subscale/scale scores were, however, negligible (*general satisfaction/clinical change* (r = -0.08; n = 92; p = 0.48) *lifestyle/ease change* (r = -0.02; n = 93; p = 0.84) and *treatment satisfaction change* (r = -0.03; n = 93; p = 0.75)). . Correlations were conducted within each ethnic group between viral load change on the one hand and the *control(c)* item and

subscale and scale scores on the other. The strength of association between *control(c)* and viral load change again differed between ethnic groups. It was significant amongst White (r = -0.37; n = 51; p = 0.007) but not African-American patients (r = -0.18; n = 32; p = 0.33). Amongst Hispanic patients, there was no association (r = 0.01; n = 9; p = 0.98). None of the correlations between viral load change and the scale and subscale scores were significant.

Discussion

The modification to the HIVTSQs has meant that a ten-item *treatment satisfaction* scale score with very good internal consistency reliability can be computed, including an important item concerning the *demands* of treatment. In addition, the two subscales, measuring *general satisfaction/clinical* aspects of treatment and satisfaction with *lifestyle/ease* of taking the treatment can also be computed, the latter including the new *demands* item. The new HIVTSQc also has a two-subscale structure, producing the *general satisfaction/clinical change* and *lifestyle/ease change* subscales, as well as a single-scale structure, producing the *treatment satisfaction change* score. All have excellent internal consistency reliability. For each measure, the subscale scores are each computed from five items and the scale score as the sum of all ten items, rendering trial analysis and interpretation of results straightforward.

The modified wording of the HIVTSQs *demands* item appears to have solved the problem identified with the earlier version. Now, all items have a similar, negatively skewed distribution. The *control* item approximated more closely a normal distribution at baseline, perhaps because patients were recruited as experiencing compromised immunity on pre-trial medication. The previous study by Woodcock and Bradley [1] found the HIVTSQs to be acceptable to patients who had been taking randomised medication for 8 or 16 weeks. In the present study, full completion at baseline by 82.9% of participants indicates that it is acceptable also to people at the start of a trial, in which some form of anti-HIV treatment was being used before the trial began. Distributions of the HIVTSQc item scores were also negatively skewed. Thus, the majority of patients found their treatment more satisfactory at Week 48 than at baseline. However, some found their trial treatment less satisfactory in certain respects, indicated by negative item scores on the scale +3 to -3. Possibly, before the trial, some

patients were not monitored as closely, nor were they given such strict instructions about medication-taking as they were for the trial. This may in part explain why the trial treatments were seen by some as presenting more *demands*, beings less *flexible*, and interfering more with *lifestyle,* so they might not recommend them *to others* as much as they did the pre-trial treatments.

The completion rate was very good for both measures. It is not known why some people did not complete the measures at all, but their demographic characteristics were similar to those of the whole sample, suggesting that non-completion may have been driven by study staff, rather than patient characteristics. Interestingly, amongst those actually attempting the HIVTSQs, African-American people were less likely to complete it in full.

In determining the structure of the HIVTSQs and c, the HIVTSQs ten-item *treatment satisfaction* scale was strongly supported (scale alpha baseline = 0.89; Week 48 = 0.91) as was the HIVTSQc ten-item *treatment satisfaction change* scale (Week 48 scale alpha = 0.92). These alphas were even higher than found for the nine-item scale in the original validation of the status version [1] (Week 8 or 16 α = 0.82 in full sample of Americans and Canadians; α = 0.80 in USA data). Moreover, the subscales for both the HIVTSQs and c had alphas >0.8, at baseline and Week 48, again higher than in the original study (*general satisfaction/clinical* α = 0.80 in USA/Canada; 0.77 in USA alone; *lifestyle/ease* α = 0.74 in USA/ Canada; 0.75 in USA). The weakest item generally, as demonstrated by standardised regression weights and Cronbach's alpha, concerned '*understanding of your HIV'.* It may be necessary to change the wording of this item in future, possibly to focus on *'understanding of your treatment',* as included in our two recently-developed measures of treatment satisfaction in paediatric diabetes, for completion by parents and by teenagers (the DTSQ-Parent and DTSQ-Teen).

Within the HIVTSQs, the revised *demands* item was the strongest item in the *treatment satisfaction* scale. Because of the high internal consistency reliability of both the *treatment satisfaction* scale and the *treatment satisfaction change* scale, up to five items may be missed from the scale, while still computing the scale score The subscales can each be computed when only one item is missed. This method of computing missing values assumes that items are missed at random and that they all measure a single underlying latent construct. In some trials using the earlier version of the HIVTSQs, the nine-item total *treatment satisfaction* scores (ranging 0-54) were converted to percentages (0-100) (Jordan et al) [8]. The same could be done with the ten-item total *treatment satisfaction* scores

(0-60), at the discretion of those using the measure. Whilst this facilitates interpretation, it tends to inflate any apparent differences between patient groups. A greater apparent inflation would result from changing HIVTSQs subscale scores (range 0-30) into percentages and when translating HIVTSQc scale scores (-30 through zero to +30) or subscale scores (-15 through zero to +15) to percentage differences from zero (-100% to +100% satisfaction). The skewed distribution of scale and subscale scores, particularly those from the change version, indicates that non-parametric statistics may be appropriate. Alternatively, ranked data may be used, or measures can be taken to normalise the skew before analysis using parametric methods. With large sample sizes, there may be very little difference between the results of non-parametric and parametric analyses, but it is advisable to check. Two years between baseline and follow-up is a long time for people to make HIVTSQc change judgments. The fact that HIVTSQc item change scores all correlated positively with the difference between status scores obtained at baseline and Week 48 provides some evidence that meaningful judgments were being made, though the HIVTSQc might best be administered after a shorter time period (6 or 12 months into a trial) to enable respondents to recall more clearly their experience of the previous treatment.

Some degree of construct validity has already been established for the HIVTSQs, by comparing treatments with more/less complex regimens [1]. The present study confirmed a relationship between viral load and *control* item scores, in both the status and change measures, and a relationship between the *general satisfaction/clinical* subscale and viral load at Week 48. In contrast, the *lifestyle/ease* subscale did not correlate with viral load at the same timepoint, indicating that the two subscales may be useful in discriminating between treatments with similar effects on clinical status, but easier/more difficult to take. Interestingly, the scale scores were not correlated with viral load (or viral load change). Thus, overall satisfaction is not simply a matter of controlling clinical status. Indeed, if viral load and satisfaction correlated strongly, there would be no need to measure satisfaction with treatment. Patients' perceptions, including convenience, and their understanding of test results can also influence satisfaction levels.

Ethnicity, viral load and treatment satisfaction

If HIVTSQs *general satisfaction/clinical, lifestyle/ease* and *treatment satisfaction* scores were directly related to clinical status, it would be expected that African-American people in this trial would be less

satisfied than Whites, because their HIV status was somewhat worse. The counterintuitive finding, that African-American people reported greater baseline satisfaction on both the subscales and the total scale, indicates that the observed ethnic group difference may be cultural rather than a reflection of different levels of disease progression. Greater treatment satisfaction amongst non-Whites was found in the earlier HIVTSQ validation [1] and contrasts with studies in which non-Whites tend to be either less satisfied with medical care than Whites [13],[14] (1996 data), [15] or equally satisfied [16],[14] (2000 data). The prevalence of non-White and particularly African-American people amongst those attending their Week 48 appointment yet missing HIVTSQc items is notable, especially because African-American people tended to report greater baseline satisfaction on the HIVTSQs. It is important to use the recommendations for dealing with missing values to compute subscale and scale scores for as many people as possible without damage to reliability at follow-up points, when drop-outs or missing data can undermine randomisation.

The correlation between viral load and HIVTSQs scores in the original study [1] was r = -0.28 with *treatment satisfaction* and r = -0.33 with the *control* item (both small but highly significant). Those patients were recruited with plasma viral load at least 400 HIV-1 RNA copies/mL and their baseline mean viral load was 4.10 copies HIV-1RNA/mL(log_{10})(sd 0.76). By the week of HIVTSQ completion (Week 8 or 16), however, their mean viral load was down to 3.10 (sd 0.85). Baseline mean viral load in the present study, transformed to a log_{10} scale, was 4.09 (sd 0.82). Thus, the viral load at the start of the two trials was very similar. At baseline of the present study, however, neither the *treatment satisfaction* scale score nor the *control* item score correlated significantly with baseline viral load. There are two possible reasons for this difference between the two studies. In the earlier study, which collected data 8 or 16 weeks into the trial, patients would by that time be familiar with viral load reporting and might have based their perceptions of HIV *control* on those readings; also, the trial protocol of the earlier study did not dictate when to administer the HIVTSQs in relation to viral load reporting and some patients may have known their results before completing the questionnaire. In the present study, however, all patients completed the HIVTSQs before receiving their test results and so a weaker correlation between HIVTSQs scores and viral load might be expected. Patients would rely on their perceptions of HIV control, which would be hard, particularly for asymptomatic patients. Viral load spanned a large range and may have included symptomatic as well as relatively symptom-free

patients. Correlations between HIVTSQs *control* item scores and viral load at baseline and Week 48 support the suggestion that participation in the trial increased understanding of viral load results, which affected ratings of the *control* item. They also suggest that African-American and particularly Hispanic participants had less information about their viral loads at baseline, possibly diluting the association between viral load and *control* item responses in the full sample. The correlation was significant amongst White participants at baseline and became significant in the combined sample by Week 48. Correlations between HIVTSQs *control* item scores and viral load at baseline and Week 48 support the suggestion that participation in the trial increased understanding of viral load results, which affected ratings of the *control* item.

For the HIVTSQc, there was no ethnic difference in subscale or scale scores, but the means indicated a tendency for African-American people to indicate the greatest improvement in satisfaction. Correlations showed that particularly White and, to a lesser extent, African-American respondents whose viral load had improved more by the close of the trial reported a greater increase in satisfaction on the HIVTSQc *control(c)* item. Despite a tendency for a similar relationship between total *treatment satisfaction change* and viral load change, the negative correlation was not significant and the correlation with the subscale scores was small in each case. The stronger correlation with *control(c)* is to be expected, because this is the one item out of the ten that is likely to relate directly to viral load and physical outcomes. This association provides preliminary evidence of construct validity of the HIVTSQc. The stronger correlation between viral load change and HIVTSQc *control(c)* scores for White patients compared with African-American or Hispanic patients suggests again that White participants might have had more information about their viral load and used this when responding to the *control(c)* item. This finding has implications for the education of people infected with HIV, but requires further investigation due to the small number of Hispanic patients in the present study.

Because of the complex relationships identified between ethnicity, understanding of viral load results and perceptions of HIV control, ethnicity could usefully be included as a potential confounding variable (covariate) in trial analyses of HIVTSQs and c data. This is in addition to ensuring that ethnic groups are similarly represented in each treatment group at recruitment, because African-Americans may also be rather more likely than other Americans to miss items. If a scale score cannot be computed for them at any point during the trial, due to missing several items, this could potentially

exclude patients indicating higher satisfaction levels from trial analyses. It is also advisable to include baseline HIVTSQs scores as a covariate to deal with any potential confounding influence on HIVTSQc scores later in the trial. The less satisfied people are when they begin the trial, the more room there will be for improvement.

These two measures are already available in several languages, produced using two independent forward translations, reconciliation, and two back translations, with revision and further back translation as needed. Further work will include assessment of the equivalence of completion methods, such as pen and paper compared with telephone administration, further linguistic validation of translations, including full cognitive debriefing with patients, as well as analysis of sensitivity to differences between treatments.

Conclusion

The HIVTSQs and c each has a structure with two five-item subscales, measuring *general satisfaction/clinical* and *lifestyle/ease* satisfaction. A ten-item *treatment satisfaction* scale may also be computed for each measure. The HIVTSQs subscale and scale scores can be used at baseline and later in the trial to make comparisons between groups and over time. The HIVTSQc subscale and scale scores allow an improvement to be expressed later in the trial, even by patients with high levels of satisfaction at baseline. The subscales and scales in both measures have very good internal consistency reliability. The *control* item score from each measure correlates with viral load more strongly than does the scale score and this relationship becomes stronger during a trial, suggesting that understanding of viral load results may improve during a trial. Together, the HIVTSQs and c can be used in trials of new HIV treatments, to provide insight into patients' perspectives, which are not as closely related to viral load as might be expected.

Acknowledgements

This work was funded by a grant from GlaxoSmithKline. We thank Maria Watson for providing trial data for analysis and for her useful comments. We also thank two anonymous reviewers and Dr Jan Mitchell for useful comments about analysis.

Figure 1. Item content of the HIVTSQs and HIVTSQc

			HIVTSQs:	**HIVTSQc**
			Instructions	**Instructions**
			The following questions are concerned with your anti-HIV medicine therapy for HIV infection and your experience over the past few weeks. Please answer each question by circling a number on each of the scales.	The following questions are concerned with your present anti-HIV medicine therapy compared with your experience of medicine therapy used just before you started in the current study. We are interested to know how, if at all, your experience of medicine therapy has changed. Please answer each question by circling a number on each of the scales to indicate the extent to which you have experienced changes. If you have experienced no change, circle '0'.
Item no.	**Item label** [Suffix *(c)* denotes item label in the change version]	**Item wording**	**Response options 6-0**	**Response options 6-0**
1	*current treatment*	How satisfied are you with your current treatment?	very satisfied 6 to 0 very dissatisfied	much more satisfied now 3 to -3 much less satisfied now
2	*control*	How well controlled do you feel your HIV has been recently?	very well controlled 6 to 0 very poorly controlled	much better controlled now 3 to -3 much worse controlled now
3	*side effects*	How satisfied are you with any side-effects of your present treatment?	very satisfied 6 to 0 very dissatisfied	much more satisfied now 3 to -3 much less satisfied now
4	*demands*	How satisfied are you with the demands made by your current treatment?	very satisfied 6 to 0 very dissatisfied	much more satisfied now 3 to -3 much less satisfied now
5	*convenience*	How convenient have you been finding your treatment to be recently?	very convenient 6 to 0 very inconvenient	much more convenient now 3 to -3 much less convenient now
6	*flexibility*	How flexible have you been finding your treatment to be recently?	very flexible 6 to 0 very inflexible	much more flexible now 3 to -3 much less flexible now
7	*understanding*	How satisfied are you with your understanding of your HIV?	very satisfied 6 to 0 very dissatisfied	much more satisfied now 3 to -3 much less satisfied now
8	*lifestyle*	How satisfied are you with the extent to which the treatment fits in with your life-style?	very satisfied 6 to 0 very dissatisfied	much more satisfied now 3 to -3 much less satisfied now
9	*recommend to others*	Would you recommend your present treatment to someone else with HIV?	Yes I would definitely recommend the treatment 6 to 0 No I would definitely not recommend the treatment	much more satisfied now 3 to -3 much less satisfied now
10	*continue*	How satisfied would you be to continue with your present form of treatment?	very satisfied 6 to 0 very dissatisfied	much more likely to recommend the treatment now 3 to -3 much less likely to recommend the treatment now
			Please make sure that you have circled one number on each of the scales.	

Table 1. Demographic characteristics, treatment group and viral load of patients included in baseline analysis of HIVTSQs data and Week 48 analysis of HIVTSQs and HIVTSQc data

	Baseline sample				Week 48 sample			
Characteristic	n	Frequency (% of sample)	Mean (sd)	Median (min-max)	n	Frequency (% of sample)	mean (sd)	median (min-max)
Age (years)	152	-	41.2 (7.9)	40 (24-69)	106	-	42.1 (7.8)	41 (27-67)
Sex:	152				106			
Men		129 (84.9%)				93 (87.7%)		
Women		23 (15.1%)				13 (12.3%)		
Ethnicity:	152							
Hispanic		14 (9.2%)				9 (8.5%)		
African-American		57 (37.5%)				39 (36.8%)		
White		80 (52.6%)				57 (53.8%)		
Other		1 (0.7%)				1 (0.9%)		
Allocated to treatment group:	152				106			
group 1		55 (36.2%)				34 (32.1%)		
group 2		47 (30.9%)				34 (32.1%)		
group 3		50 (32.9%)				38 (35.8%)		
Viral load (HIV-1 RNA copies/mL)	141		56023.1 (115960.7)	10915 (49-606043)	105		21181.7 (90724.8)	49 (49-671220)
<1000		12 (8.5%)				83 (79.0%)		
≥1000		129 (91.5%)				22 (21.0%)		
HIV-1 RNA copies/ml(log_{10})	141		4.09 (0.82)	4.04 (1.69-5.78)	105		2.34 (1.10)	1.69 (1.69-5.83)

Table 2. Confirmatory Factor Analysis (CFA) of the ten HIVTSQs items (baseline and Week 48).

(showing standardised regression weights for each variable and normed fit index (NFI) for the model.)

Item	**Baseline**		**Baseline**	**Week 48**		**Week 48**
	2 latent variables		**Single latent variable**	**2 latent variables**		**Single latent variable**
	General satisfaction/clinical subscale	*Lifestyle/ease* subscale	*Total treatment satisfaction* scale	*General satisfaction/clinical* subscale	*Lifestyle/ease* subscale	*Total treatment satisfaction* scale
	NFI = 0.97		NFI = 0.95	NFI =0.98		NFI = 0.97
1. *current treatment*	0.846		0.725	0.914		0.798
2. *control*	0.767		0.671	0.767		0.665
3. *side effects*	0.514		0.595	0.800		0.693
4. *demands*		0.853	0.861		0.879	0.876
5. *convenience*		0.894	0.812		0.950	0.923
6. *flexibility*		0.654	0.632		0.669	0.675
7. *understanding*		0.378	0.368		0.409	0.400
8. *lifestyle*		0.720	0.661		0.882	0.859
9. *recommend to others*	0.634		0.693	0.693		0.705
10. *continue*	0.820		0.698	0.742		0.734

Table 3. Internal consistency reliability of HIVTSQs subscales and scale (baseline).

Item	*General satisfaction/clinical subscale* scale alpha = 0.831		*Lifestyle/ease subscale* scale alpha =0.821		*Total treatment satisfaction scale* scale alpha = 0.891	
	Corrected item-total correlation	Cronbach's α if item deleted	Corrected item-total correlation	Cronbach's α if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
1. *current treatment*	0.74	0.767			0.68	0.876
2. *control*	0.67	0.788			0.67	0.877
3. *side effects*	0.46	0.845			0.54	0.887
4. *demands*			0.66	0.775	0.79	0.868
5. *convenience*			0.77	0.742	0.73	0.876
6. *flexibility*			0.55	0.812	0.56	0.885
7. *understanding*			0.43	0.834	0.40	0.893
8. *lifestyle*			0.71	0.758	0.66	0.878
9. *recommend to others*	0.60	0.806			0.69	0.876
10 *continue*	0.71	0.774			0.65	0.879

Table 4. HIVTSQs subscale and scale scores of demographic subgroups (baseline)

Characteristic (for test used, see footnote [T])	Subgroups of patients compared (n indicates number in subgroup)HIVTSQs subscale/scale score with Mean (sd) Median (min-max)				Significance
General satisfaction/clinical subscale					
Age category	Younger (n=66) Mean 20.1 (6.8) Median 21 (2-30)	Older (n=70) Mean 21.2 (6.7) Median 22 (4-30)			p = 0.33
Sex	Men (n=116) Mean 20.4(6.7) Median 21.5(2-30)	Women (n=20) Mean 22.0(6.7) Median 22.3(4-30)			p = 0.29
Ethnicity[¥]	Hispanic (n=13) Mean 22.5(6.5) Median 23 (10-29)	African-American (n=50) Mean 22.0 (6.9) Median 23 (4-30)	White (n=72) Mean 19.4 (6.5) Median 20.5 (2-30)		p = 0.03
	Hispanic	African-American			p = 0.98
	Hispanic		White		p = 0.09
		African-American	White		p = 0.02
Lifestyle/ease subscale					
Age category	Younger (n=66) Mean 23.1 (5.4) Median 24 (8-30)	Older (n=71) Mean 22.6 (5.8) Median 24 (1-30)			p = 0.71
Sex	Men (n=117) Mean 23.1 (5.4) Median 24 (8-30)	Women (n=20) Mean 21.6(6.8) Median 23.0(1-29)			p = 0.33
Ethnicity[¥]	Hispanic (n=13) Mean 24.4 (6.1) Median 27 (11.25-30)	African-American (n=50) Mean 24.1 (5.6) Median 25 (1-30)	White (n=73) Mean 21.9 (5.3) Median 23 (8-30)		p = 0.01
	Hispanic	African-American			p = 0.53
	Hispanic		White		p = 0.08
		African-American	White		p = 0.006
Total satisfaction scale					
Age category	Younger (n= 68) Mean 43.2 (10.8) Median 11.0 (17-60)	Older (n= 71) Mean 43.7 (11.76) Median 16.0 (5-60)			p = 0.64
Sex	Men (n=119) Mean 43.4 (11.1) Median 45.0 (16-60)	Women (n=20) Mean 43.6 (12.4) Median 44.5 (5-58.9)			p = 0.86
Ethnicity[¥]	Hispanic (n=13) Mean 47.0 (11.0) Median 52.0(27-58)	African-American (n= 52) Mean 46.0 (11.53) Median 47.5 (5-60)	White (n= 73) Mean 41.2 (10.7) Median 43.0(17-60)		p = 0.014
	Hispanic	African-American			p = 0.90
	Hispanic		White		p = 0.07
		African-American	White		p = 0.007

[T] Mann-Whitney to compare two age categories (median split at 40 years) and to compare men and women. Kruskal-Wallis (with χ^2) to compare three ethic groups, followed by Mann-Whitney between pairs of ethnic group. [¥] Other' ethnicity (n=1) not included in analyses, due to insufficient number of cases.

Table 5. Confirmatory Factor Analysis (CFA) of the ten HIVTSQc items (Week 48)
(showing standardised regression weights for each variable and normed fit index (NFI) for the model.)

Item	2 latent variables		Single latent variable
	General satisfaction/clinical change subscale	*Lifestyle/ease change subscale*	*Total treatment satisfaction change scale*
	NFI = 0.92		NFI = 0.90
1. current treatment(c)	0.863		0.815
2. control(c)	0.403		0.437
3. side effects(c)	0.611		0.649
4. demands(c)		0.832	0.841
5. convenience(c)		0.821	0.761
6. flexibility(c)		0.819	0.795
7. understanding(c)		0.478	0.480
8. lifestyle(c)		0.951	0.912
9. recommend to others(c)	0.759		0.686
10. continue(c)	0.981		0.903

Table 6. Internal consistency reliability of the HIVTSQc subscales and scale (Week 48).

Item	*General satisfaction/clinical* subscale scale alpha =0.850		*Lifestyle/ease* subscale scale alpha =0.882		*Total satisfaction change* scale scale alpha = 0.916	
	Corrected item-total correlation	Cronbach's α if item deleted	Corrected item-total correlation	Cronbach's α if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
1. current treatment(c)	0.83	0.78			0.77	0.903
2. control(c)	0.40	0.88			0.38	0.921
3. side effects(c)	0.59	0.84			0.59	0.912
4. demands(c)			0.73	0.859	0.79	0.903
5. convenience(c)			0.74	0.852	0.69	0.907
6. flexibility(c)			0.82	0.830	0.77	0.903
7. understanding(c)			0.50	0.904	0.52	0.917
8. lifestyle(c)			0.84	0.824	0.84	0.897
9. recommend to others(c)	0.67	0.92			0.71	0.906
10. continue(c)	0.84	0.77			0.84	0.898

Table 7. HIVTSQc *treatment satisfaction change* scores of demographic subgroups (Week 48)

Characteristic (for test used, see footnote [T])	Subgroups of patients compared, with number in subgroup (n) HIVTSQc subscale/scale score Mean (sd) and Median (min to max)			Significance
General satisfaction/clinical change subscale				
Age category	Younger (n=40) Mean 12.8 (2.3) Median 14 (6 to 15)	Older (n=59) Mean 11.2 (5.2) Median 13 (-5 to 15)		p = 0.44
Sex	Men (n=88) Mean 11.7 (4.5) Median 13 (-5 to 15)	Women (n=11) Mean 13.3 (1.8) Median 14 (10 to15)		p = 0.41
Ethnicity[¥]	Hispanic (n=9) Mean 9.9(7.8) Median 14 (-5 to15)	African-American (n=33) Mean 12.3 (4.0) Median 14 (-4 to 15)	White (n=56) Mean 11.9 (3.7) Median 13 (-3 to 15)	p = 0.56
Lifestyle/ease change subscale				
Age category	Younger (n=40) Mean 10.7 (4.4) Median 11 (-6 to 15)	Older (n=60) Mean 10.9 (5.1) Median 13 (-4 to 15)		p = 0.51
Sex	Men (n=89) Mean 10.7 (4.9) Median 12 (-6 to 15)	Women (n=11) Mean 11.5 (4.1) Median 13.8 (4 to 15)		p = 0.56
Ethnicity[¥]	Hispanic (n=9) Mean 10.1 (6.3) Median 14 (-2.5 to15)	African-American (n=34) Mean 11.8 (4.6) Median 14 (-4 to 15)	White (n=56) Mean 10.4 (4.6) Median 11 (-6 to 15)	p = 0.22
Treatment satisfaction change scale				
Age category	Younger (n=40) Mean 23.5 (5.9) Median 24 (6 to 30)	Older (n=60) Mean 22.0 (10.0) Median 26 (-8.26 to 30)		p = 0.81
Sex	Men (n=89) Mean 22.3 (8.9) Median 26 (-8.26 to 30)	Women (n=11) Mean 24.8 (5.4) Median 27.8 (17 to 30)		p = 0.47
Ethnicity[¥]	Hispanic (n=9) Mean 19.9 (14.4) Median 27 (-8.26 to 30)	African-American (n=34) Mean 24.0 (8.3) Median 27.9 (-8 to 30)	White (n=56) Mean 22.3 (7.6) Median 23.5 (-3 to 30)	p = 0.25

[T] Mann-Whitney U to compare two age categories (with median split at 40 years) and to compare men and women. Kruskal-Wallis (with χ^2) to compare three ethic groups.[¥] 'Other' ethnicity (n=1) not included in analysis due to insufficient number of cases. Post-hoc Mann Whitney not conducted if Kruskal-Wallis not significant

Access to Questionnaires

The HIVTSQs and c are available from the copyright holder, Prof Clare Bradley: c.bradley@rhul.ac.uk

References

[1] Woodcock A, Bradley C. Validation of the HIV Treatment Satisfaction Questionnaire. Quality of Life Research 2001; 10: 517-531.

[2]Bradley C, Lewis KS (1990) Measures of psychological well-being and treatment satisfaction developed from the responses of people with tablet-treated diabetes. Diabetic Medicine 7, 445-451.

[3]Bradley C. The Diabetes Treatment Satisfaction Questionnaire (DTSQ) Chapter 7 pp111-132 In Bradley Ed Handbook of Psychology and Diabetes: a guide to psychological measurement in diabetes research and practice Chur Switzerland; Harwood Academic. 1994.

[4] Barendse SM, Speight J, Bradley C. The Renal Treatment Satisfaction Questionnaire (RTSQ): a measure of satisfaction with treatment for chronic kidney failure. Am J Kidney Dis. 2005; 45 (3):572-9.

[5] Taback N, Bradley C Validation of the Genital Herpes Treatment Satisfaction Questionnaire (GHerpTSQ) in status and change versions. Quality of Life Research, in press 2006.

[6]Woodcock A, Plowright R, Kennedy-Martin T, Hirsch A, ffytche T, Bradley C. (2005) Development of the new Retinopathy Treatment Satisfaction Questionnaire (RetTSQ) *International Congress Series* 1282: 342-346.

[7] McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU: Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. Thyroid 2004, 14(11):916-925.

[8] Jordan J, Cahn P, Goebel F, Matheron S, Bradley C, Woodcock A. Abacavir Compared to Protease Inhibitors as Part of Highly Active Antiretroviral (HAART) Regimens for Treatment of HIV Infection: Better Patient Satisfaction and Implications for Adherence AIDS Patient Care and STDs 2005; 19 (1) 9-18.

[9] Bradley C. Diabetes Treatment Satisfaction Questionnaire: Change version for use alongside status version provides appropriate solution where ceiling effects occur. Diabetes Care 1999; 22: 530-532

[10] Bradley C, Plowright R, Stewart J, Witthaus E. Diabetes Treatment Satisfaction Questionnaire (change) in English and German, evaluated in insulin glargine trials. Diabetologia 2000; 43: (suppl 1) A196 (754).

[11] Howorka K, Pumprla J, Schlusche C, Wagner-Nosiska D, Schabmann A, Bradley C. Dealing with Ceiling Baseline Treatment Satisfaction Level in Patients with Diabetes under Flexible, Functional Insulin Treatment: Assessment of Improvements in Treatment Satisfaction with a New Insulin Analogue. Quality of Life Research 2000; 9: 915-930.

[12] Arbuckle JL, Wothle W. Amos 4.0 User's Guide. Chicago: Smallwaters Corporation 1999.

[13] Lantz PM, Janz NK, Fagerlin A, Schwartz K, Liu L, Lakhani I, Salem B, Katz SJ. Satisfaction with surgery outcomes and the decision process in a population-based sample of women with breast cancer. Health Services Research 2005; 40:745-67.

[14] Shi L, Stevens GD. Disparities in access to care and satisfaction among U.S. children: the roles of race/ethnicity and poverty status. Public Health Reports 2005; 120: 431-41.

[15] Woods SE, Bivins R, Oteng K, Engel A. The influence of ethnicity on patient satisfaction. Ethnicity and Health 2005; 10: 235-42.

[16] De Wilde EF, Hendriks VM. The Client Satisfaction Questionnaire: psychometric properties in a Dutch addict population. European Addiction Research 2005; 11:157-62.