BMC Medical Research Methodology

BioMed Central

Research article

Open Access

# Methodological standards in non-inferiority AIDS trials: moving from adherence to compliance

Jean-Jacques Parienti*[1,2], Renaud Verdon[3] and Véronique Massari[1]

Address: [1]Inserm UMR-S 707, Paris, F-75012; Université Pierre et Marie Curie-Paris6, UMR-S 707, Paris, F-75012, France, [2]Department of Biostatistics and Clinical Research, Côte de Nacre University hospital, 14033 Caen, France and [3]Department of Infectious Diseases, Côte de Nacre University hospital, 14033 Caen, France

Email: Jean-Jacques Parienti* - parienti-jj@chu-caen.fr; Renaud Verdon - verdon-r@chu-caen.fr; Véronique Massari - massari@u707.jussieu.fr

* Corresponding author

Published: 20 September 2006

*BMC Medical Research Methodology* 2006, **6**:46 doi:10.1186/1471-2288-6-46

Received: 03 January 2006

Accepted: 20 September 2006

This article is available from: http://www.biomedcentral.com/1471-2288/6/46

© 2006 Parienti et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** The interpretation of the results of active-control trials regarding the efficacy and safety of a new drug is important for drug registration and following clinical use. It has been suggested that non-inferiority and equivalence studies are not reported with the same quantitative rigor as superiority studies.

**Methods:** Standard methodological criteria for non-inferiority and equivalence trials including design, analysis and interpretation issues were applied to 18 recently conducted large non-inferiority (15) and equivalence (3) randomized trials in the field of AIDS antiretroviral therapy. We used the continuity-corrected non-inferiority chi-square to test 95% confidence interval treatment difference against the predefined non-inferiority margin.

**Results:** The pre-specified non-inferiority margin ranged from 10% to 15%. Only 4 studies provided justification for their choice. 39% of the studies (7/18) reported only intent-to-treat (ITT) analysis for the primary endpoint. When on-treatment (OT) and ITT statistical analyses were provided, ITT was favoured over OT for results interpretation for all but one study, inappropriately in this statistical context. All but two of the studies concluded there was "similar" efficacy of the experimental group. However, 9/18 had inconclusive results for non-inferiority.

**Conclusion:** Conclusions about non-inferiority should be drawn on the basis of the confidence interval analysis of an appropriate primary endpoint, using the predefined criteria for non-inferiority, in both OT and ITT, in compliance with the non-inferiority and equivalence CONSORT statement. We suggest that the use of the non-inferiority chi-square test may provide additional useful information.

## Background

Equivalence and non-inferiority randomized controlled trials are the standard research methodology to demonstrate that a new treatment is equivalent or non-inferior to standard therapy (active-control) in term of efficacy. While an equivalence trial would use the 2-sided 95% confidence interval of the difference between the 2 trial arms, the non-inferiority trial would usually use the 90% confidence interval of the difference, if a 1-sided 5% rather than 2.5% significance test was considered a priori acceptable [1]. Because it is impossible to prove exact equality, the goal in a non-inferiority trial, in situations

where the effect compared to placebo is large, is to rule out differences of clinical importance in the primary outcome between the two treatments.

Issues, difficulties and controversies surrounding non-inferiority trials have long been well recognized and extensively reported in many medical settings, including human immunodeficiency virus infection (HIV) [2,3]. Highly active antiretroviral therapy (HAART) delays progression of the acquired immunodeficiency syndrome (AIDS) and increases survival among HIV infected patients. With efficacy rates of 70% [4] and 75% [5] respectively, the space for better antiretroviral agents efficacy has become very tight. However, long term toxicities, pill burden and genotypic resistance call for treatment simplification and alternative new agents. As a consequence, the number of non-inferiority trials has been growing in the recent years in the AIDS therapy literature. Some authors chose to use interchangeably the terms "equivalence" and "non-inferiority", regardless of the hypothesis of the study. Given that the question of interest is not symmetric, we think that they are better described as "non-inferiority" trials[6].

Because efficacy in viral suppression remains the major outcome, new drugs should first prove non-inferiority with respect to prolonged control of HIV replication, as the primary endpoint. Second, the new drugs should provide other advantages. Inevitably, there may have been some tension between marketing purposes and scientific issues in the published reports of those trials. In this paper, our objective was to verify the validity of recently published non-inferiority AIDS trials regarding the primary endpoint.

## Methods

### *Study selection and methodological standards*

Our aim was to consider a cohort of equivalence or non-inferiority trials published in the area of HIV/AIDS, after HAART became available. We performed a MEDLINE search using the terms equivalence OR non-inferiority AND random* AND HIV (1) and abacavir AND random* (2). 64 (1) and 136 (2) articles were identified. 5 (1) and 5 (2) were selected because they fulfilled the following requirements: randomized controlled clinical trial with 48-week minimum follow-up, initially designed as a non-inferiority or equivalence trial with a prespecified non-inferiority margin, virological primary endpoint and publication in *New England Journal Medicine*, *JAMA*, *Lancet*, *AIDS*, *Clinical Infectious Diseases*, *Journal of Infectious Diseases and Journal of Acquired Immune Deficiency Syndrome* between 2001 and 2006. Eight additional articles were identified by examining cross-references or by authors' knowledge of their existence.

We applied traditional methodological requirements for non-inferiority and equivalence trials adapted from Kirshner[7], Jones et al. [8], McAlister and Sackett [9] and Piaggio et al.[1] to eighteen [10-27] active-control trials. We also applied proposed standards in the report of non-inferiority and equivalence trials adapted from Le Henanff [28].

### *Statistical analysis*

Intent-to-treat (ITT) or on-treatment (OT) analysis 95% confidence interval of the treatment difference were computed using the normal approximation, based on available data included in the flow chart, results section and figures. Two selected studies (ALIZE and SEAL) predefined a 90% confidence interval of the treatment difference, but their conclusions were not affected by the use of the 95% confidence interval (which was used in this paper for homogeneity). Two other selected studies (BMS-045 and CONTEXT) defined the primary endpoint as the $\log_{10}$ reduction in HIV viral load, using a time-averaged difference method. For homogeneity with other studies, we considered the more pertinent criteria (closer to the clinical practice) of the percentage of patients with undetectable viral load (< 50 copies/ml or < 400 copies/ml) at week 48 (reported as secondary endpoint).

In case of missing data, the corresponding author of the paper was contacted. When only percentages were available with several possibilities for the numerator due to rounding, we choose on a worst case basis. If original data were censored, we used the cumulative incidence of the primary endpoint in each arm.

Significance testing in establishing non-inferiority between the two arms of a study was computed by the use of the continuity-corrected chi-square of Dunnett and Gent [29] for non-inferiority in intent-to-treat or on-treatment analysis, also on a worst case basis. Briefly, $\pi_1$ and $\pi_2$ represent the true proportions of patients with treatment success according to the primary outcome in a random sample of the 2 populations of patients receiving the control treatment and the new drug, respectively. In case of non-inferiority, the expected estimates of $\pi_1$ and $\pi_2$ are given by:

$$\hat{\pi}_1 = \frac{x + \gamma + n_2\Delta}{n_1 + n_2} \quad \hat{\pi}_2 = \frac{x + \gamma - n_1\Delta}{n_1 + n_2}$$

where $x$ and $\gamma$ are the observed number of success, $n_1$ and $n_2$ are the 2 sample sizes in the control and the experimental study groups, respectively and $\Delta$ the pre-specified margin for non-inferiority.

The continuity-corrected chi-square of Dunnett and Gent [29] (reproduced with written permission) for non-inferiority is given by:

$$\chi_c^2 = (|x - x| - \frac{1}{2})^2 \left[ \frac{1}{x} + \frac{1}{m-x} + \frac{1}{n_1 - x} + \frac{1}{n_2 - m + x} \right]$$

where $m = x + y$ and $\hat{x} = \hat{\pi}_1 n_1$

If Δ is the maximal acceptable difference in success rates between the 2 treatment arms and δ is the observed difference between the experimental and control arms, the equivalence hypothesis can be formulated as pair of one-sided hypothesis:

$H_{01}: \delta \geq \Delta$ versus $H_{a1}: \delta < \Delta$ with a type I error of $\alpha_1$ (1)

and

$H_{02}: \delta \geq -\Delta$ versus $H_{a2}: \delta > -\Delta$ with a type I error of $\alpha_2$ (2)

The type I error probability α for $H_0$ rejection corresponds to $H_{01} \cup H_{02}$. Therefore, the P-value for equivalence is the lower chi-square value associated with max ($\alpha_1$, $\alpha_2$). In a non-inferiority hypothesis, only (1) is necessary. More details have been published elsewhere[30].

To avoid confusion between the P-values of superiority tests and the P-values of non-inferiority tests (both are reported in this paper), the latter have been renamed "D-values". When the normal approximation is a valid hypothesis, there is a general consistency between the two-sided 95% confidence interval approach (non-inferiority at $\alpha/2 < 2.5\%$) and the non-inferiority chi-square (D-value < 5%), as shown in Figure 1. D-values and P-values < 0.05 were considered statistically significant.

## Results

### *Efficacy of the active control and similar outcome*

All of the antiretroviral trials outlined in Table 1 were conducted with active-controls which have previously shown efficacy. 16 studies used a composite endpoint including virologic failure, clinical progression to AIDS or death in compliance with the other new AIDS clinical trials, whereas 2 studies used $\log_{10}$ reduction in HIV viral load. However, they reported virologic failure as secondary endpoints.

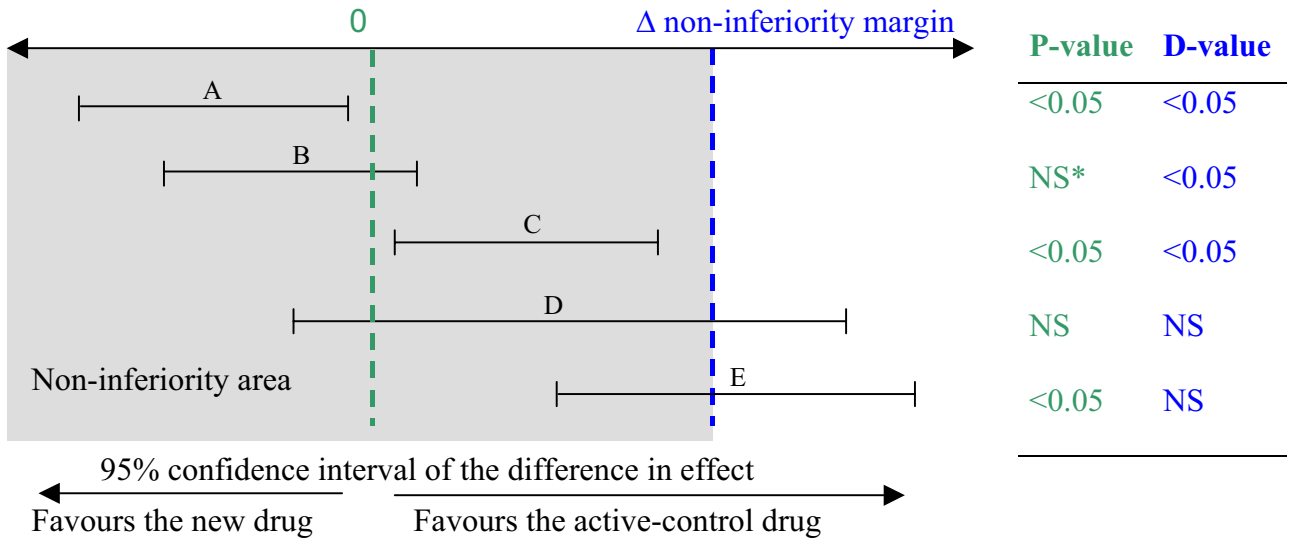


**Figure 1**

Correspondence between 95% confidence interval of the difference in effect, superiority P-value and non-inferiority D-value. * NS indicates non-significance for superiority or non-inferiority. Case A shows significant superiority of the new drug and necessarily non-inferiority Case B shows significant non-inferiority, but superiority of the new drug is uncertain (inconclusive result) Case C shows both, significant inferiority of the new drug (or superiority of active-control) but nonetheless significant non-inferiority Cases D and E failed to demonstrate non-inferiority (inconclusive result) but E demonstrated significant inferiority (or superiority of active-control).

**Table 1: Characteristics of the 18 non-inferiority studies**

| | Power (%) | Double blind | Experimental arm(s) (ITT sample size) | Control arm (ITT sample size) |
|---|---|---|---|---|
| CNAAB3005 [10] | NA | Yes | ABA (262) | PI-based regimen (265) |
| NEFA [11] | 90 | No | (a) NVP BID (155)<br>(b) ABA BID (149) | EFV QD (156) |
| BEST [12] | 90 | No | IDV/RITO BID (162) | IDV TID (161) |
| 2NN [13] | 80 | No | (a) NVP QD (220)<br>(b) NVP BID (387)<br>(c) NVP+EFV (209) | EFV (400) |
| 903 [14] | 80 | Yes | TNF (299) | Stavudine (301) |
| SOLO [15] | 85 | No | FPV/RITO QD (322) | Nelfinavir BID (327) |
| FTC-303 [16] | 85 | No | FTC QD (294) | 3TC BID (146) |
| EPV20001 [17] | 80 | Yes | 3TC QD (278) | 3TC BID (276) |
| ALIZE [18] | 80 | No | FTC-ddI-EFV QD (178) | PI-based regimen BID/TID (177) |
| CNA30024 [19] | 85 | Yes | ABA BID (324) | AZT BID(325) |
| BMS-2004 [20] | 90 | Yes | Atazanavir (405) | EFV (405) |
| ESS40013 [21] | 80 | No | Stop EFV (141) | Continue EFV (141) |
| SEAL [22] | 80 | No | 3TC+ABA QD (130) | 3TC+ABA BID (130) |
| BMS-045 [23] | NA | No | (a) ATA/RITO QD (120)<br>(b) ATA/SAQUI (115) | LOPI/RITO BID (123) |
| CNA30021 [24] | 90 | Yes | ABA QD (384) | ABA BID (386) |
| CONTEXT [25] | NA | No | (a) FPV/RITO QD (105)<br>(b) FPV/RITO BID (107) | LOPI/RITO BID (103) |
| SHAART [26] | 80 | No | ABA BID (68) | NVP BID (66) |
| 934 [27] | 85 | No | TNF+FTC QD (255) | AZT+3TC BID (259) |

*one-sided 5% type I error
** one-sided 1.25% type I error
Abbreviations : NA: not available; E: equivalence; NI: non-inferiority; ABA: abacavir PI: protease inhibitor ; NVP: nevirapine ; EFV: efavirenz ; IDV: indinavir ; RITO: ritonavir ; TNF : tenofovir ; FPV: fosamprenavir; FTC : emtricitabine; 3TC: lamivudine ; ddI: didanosine ; AZT : zidovudine ; QD: once-a-day ; BID: twice-a-day ; TID three times-a-day

### *Rationale for the non-inferiority margin*

All studies identified a pre-specified non-inferiority margin (criterion for selection). As shown in Figure 2, however, only 4/18 studies reported justification for their choice. In the CNAAB3005 study, the choice of the non-inferiority margin was based on discussion with clinical investigators and with the Food and Drug Administration. The margin of 12% was considered as the largest difference clinically acceptable. In the 903 study, the authors considered that the margin of 10% was a more stringent and conservative non-inferiority criterion. The authors of the CNA30024 commented that it was the appropriate measure for distinguishing the clinical effectiveness of 2 study treatment. Finally, the CNA30024 authors' choice relied on HIV clinicians' judgement as well as on discussion with independent reviewers. Other studies did not comment on their choice, which ranged from 10% to 15% (median: 12%). CONTEXT and BMS-045 considered a non-inferiority margin of -0.5 $\log_{10}$ reduction in HIV viral load, without justification. Other issues regarding design are reported in Table 1.

### *Confidence interval and superiority testing*

All but two trials reported results using the confidence interval approach. In the BEST study, the authors predefined their non-inferiority margin for sample size calculation, but the confidence interval was neither defined nor reported. In the NEFA study, although the confidence interval approach was clearly defined in the statistical analysis section of the article, none was provided in the results section. NEFA, BEST, 2NN, FTC-303, ESS40013 and SHAART studies reported non-significant superiority tests for efficacy to reinforce non-inferiority. The ALIZE and 934 studies switched from the non-inferiority to the superiority hypothesis to declare that the experimental treatment had superior efficacy in the ITT analysis set (for secondary and primary endpoints, respectively), as appropriate.

### *Intent-to-treat and on-treatment analysis on the primary endpoint*

CNAAB3005, NEFA, SOLO, BEST, EPV20001, ALIZE, BMS-2004, SEAL and SHAART studies (Figure 2) published both ITT and OT analysis (9/18), but only the ALIZE, SOLO, EPV20001 BMS-2004 and SEAL studies found concordant results regarding non-inferiority in the two analysis. The BEST investigators provided separate conclusions for ITT and OT, as appropriate. The ALIZE-trial group conducted ITT, OT and a worst scenario analysis. In CNAAB3005 NEFA and SHAART, the conclusion was based on ITT analysis only. 2NN, FTC-303, EPV20001, ESS40013, CNA30021 and 934 studies

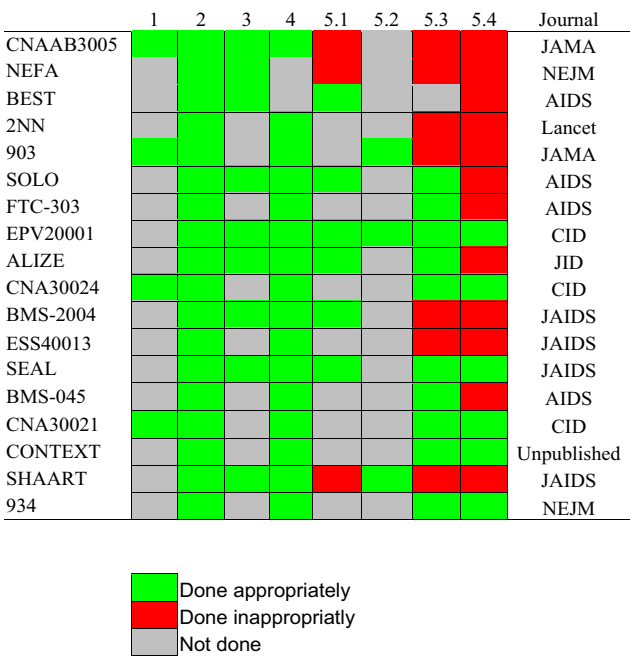


**Figure 2**

Quality report assessment of non-inferiority trials adapted from Le Henanff et al. [28]. 1. Report the margin and the justification for its choice 2. Appropriate sample size calculation 3. Report both on-treatment and intent-to-treat analysis for the primary endpoint 4. Report 1-sided or 2-sided confidence intervals of treatment difference 5. Conclusion 5.1 Conclude non-inferiority or equivalence only if both ITT and OT analyses permit that or provide separate conclusions. 5.2. Restate the prespecified margin in the abstract 5.3 Make interpretation according to the margin of equivalence or non-inferiority regarding of the primary endpoint 5.4 Conclude with standard and appropriate vocabulary in accordance with the aim and the results of the trial (ie, "non-inferior to" or "equivalent to").

described sufficient details to permit alternative analyses, such as OT. We have failed to compute OT analysis from the 903 and CNA30024 studies. Because of the nature of their primary outcome, CONTEXT and BMS-045 studies were not able to provide ITT and OT analysis. Both analysis were provided as secondary endpoints.

#### *Interpretation and conclusion (non-inferiority margin versus observed upper bound of the 95% confidence interval of the difference)*

CNAAB3005 (12% versus 14.3), NEFA (13.5% versus 15.8), 2NN (10% versus 14.0%; 10% versus 14.6%), 903 (10% versus 10.3%) and SHAART (15% versus 17.4%) concluded non-inferiority inappropriately on the basis of their pre-specified margin. In accordance, their non-inferiority D-values were above 5%, as shown in Table 2. BMS-2004 concluded that the two drugs were as efficacious (suggesting equivalence), while the ITT lower bound of the 95% confidence interval (-11.7%) exceeded 10% in favour of the experimental drug. The main BMS-2004 hypothesis (non-inferiority of the experimental drug at 10%) was demonstrated with a D-value = 0.043 (OT analysis). In our analysis of the ESS40013 study (OT), thenon-inferiority margin exceeded the pre-specified non-inferiority margin. Finally, CONTEXT and BMS-045 studies provided a conclusion in accordance with their non-inferiority margin (data not shown).

BEST, SOLO, FTC-303, EPV20001, ALIZE, CNA30024, SEAL, CNA30021 and 934 conclusions' were appropriate, on the basis of available data.

## Discussion

Trials that assess non-inferiority require rigorous methods for their design, analysis and interpretation. Although the design and the sample size were appropriate for AIDS non-inferiority and equivalence trials, there is room for substantial improvement regarding statistical analysis and interpretation of the results.

Patients with HIV infection would be harmed by deferral of therapy. Consequently, the use of placebo would be unethical [2]. Even if placebo-controlled of HAART therapy are not available, a conclusion about efficacy can be reached because the great majority of patients (about 70%) will not be controlled without treatment [4,5]. Because significant inferiority to active-control would be a major problem for patients, the non-inferiority margin for a new drug should be smaller than the difference between active-control and placebo. Because this effect size is so large, only the clinically chosen margin is really an issue, but is also highly subjective. As a result, this margin varied from the conventional 10% up to 15%. Even the same study group chose different margins in studies 903 (10%) and 934 (13%). A small decrease in margin provides greater assurance of satisfactory effect, but the cost of the study will increase because more patients are required. In the 903-study, the authors could not demonstrate non-inferiority at 10% but they point out in their discussion that this margin was more stringent than the 12% chosen in CNAAB4005. However, if the authors had chosen the less powerful 12% as the maximal limit for non-inferiority, the 95% confidence interval would have been wider, possibly beyond the 12% limit. Consequently, data-driven discussion about the non-inferiority margin after completion of the study is pointless.

Blinding has been described as less efficient in non-inferiority than superiority trials, in particular if the primary endpoint is subjective[31]. For example, a blinded investigator could bias the results toward a preconceived belief

**Table 2: Results of the 18 non-inferiority studies**

| | Δ Maximum difference % (hypothesis) | δ (%)* | UBCI§ of δ (%) | Non-inferiority D-value (type of analysis) | Superiority P-value (ITT) | Abstract authors' conclusion as compared with active-control for primary endpoint |
|---|---|---|---|---|---|---|
| CNAAB3005 | 12 (E) | 7.3 | **14.3** | 0.17 (OT) | 0.90 | Was equivalent to |
| NEFA | 13.5 (NI) | | | | | |
| (a) | | 3.9 | 9.8 | < 0.001 (ITT) | 0.20 | No conclusion |
| (b) | | 8.5 | **15.8** | 0.14 (OT) | 0.035 | A trend toward higher rate of [failure] |
| BEST | 15 (NI) | 15.7 (ITT) | **25.9** (ITT) | 1.0 (ITT) | 0.003 | Superiority of control in ITT |
| | | 0.7 (OT) | 7.2 (OT) | < 0.001 (OT) | 1.0 (OT) | Non-inferiority in OT |
| 2NN | 10 (E) | | | | | |
| (a) | | 5.9 | **14.0** | 0.36 (ITT) | 0.15 | Showed similar efficacy |
| (b) | | 7.7 | **14.6** | 0.56 (OT) | 0.091** | Showed similar efficacy |
| (c) | | 15.4 | **23.6** | 0.24 (ITT) | 0.0003 | Did not show efficacy |
| 903 | 10 (NI) | 4.1 | **10.3** | 0.07 (ITT) | 0.19 | Highly effective and comparable*** |
| SOLO | 12 (NI) | -1.0 | 6.1 | < 0.001 (ITT) | 0.78 | Provided durable [efficacy] |
| FTC-303 | 15 (NI) | 5.0 | 12.9 | < 0.01 (ITT) | 0.234 | Was equivalent to |
| EPV20001 | 12 (E) | -0.1 | 3.6 | < 0.001 (OT) | 0.81 | Regimens were equivalent |
| ALIZE | 15 (NI*) | -2.9 | 3.6 | < 0.001 (ITT) | 0.39 | Associated with sustained [efficacy] |
| CNA30024 | 12 (NI) | -0.8 | -6.3 | < 0.001 (ITT) | 0.82 | Not inferior to |
| BMS-2004 | 10 (NI) | 3.9 | 9.7 | 0.043 (OT) | 0.16 | As efficacious as |
| ESS40013 | 12 (NI) | 6.1 | **13.5** | 0.09 (OT) | 0.77 | Maintained [efficacy] |
| SEAL | 12 (NI*) | -1.5 | 4.3 | < 0.001 (ITT) | 0.61 | Not inferior to |
| BMS-045 | NA (NI) | | | | | |
| (a) | | 8.0*** | 20.4*** | NA (ITT) | 0.21 | As effective as |
| (b) | | 19.5*** | 32.2*** | NA (ITT) | 0.003 | Efficacy was lower than |
| CNA30021 | 12 (NI) | 2.2 | 6.6 | < 0.001 (OT) | 0.61 | Not inferior to |
| CONTEXT | NA (NI) | | | | | |
| (a) | | 14.0*** | 28.0*** | NA (OT) | 0.07 | Not shown to be as effective as |
| (b) | | 10.0*** | 23.0*** | NA (ITT) | 0.30 | Not shown to be as effective as |
| SHAART | 15 (NI) | 2.1 | **17.4** | 0.13 (ITT) | 0.784 | Not inferior to |
| 934 | 13 (NI) | -3.5 | 0.5 | < 0.001 (OT) | < 0.005 | Fulfilled criteria for non-inferiority and proved superior |

§ Bold UBCI exceeded the pre-specified non-inferiority margin

*A positive δ corresponds to a higher efficacy of the active-control group, as compared with the experimental group. We choose the OT or ITT in a worst case basis, unless the authors reached separate conclusions for OT and ITT. The numbers may differ from original reports because original reports were stratum-adjusted or used 90% confidence interval.

** If patients who never started treatment were excluded, P = 0.03

***Based on secondary endpoints

Abbreviations: δ : Observed difference between the % of success observed in the control arm minus the % of success observed in the experimental arm ; UBCI : upper bound of the 95% confidence interval; E: equivalence; NI: non-inferiority

in equivalence by assigning similar ratings to the treatment responses of all patients, giving a "bias toward the null". Even when the primary outcome is objective (viral failure, clinical progression or death), however, we believe that blinding is important to protect against bias. Unblinded investigators may provide other effective therapies to patients in the arm that they believe superior or equivalent, such as more regular appointment or adherence support. In addition, patient or physicians may over-interpret subjective endpoints such as side-effects in open-label studies. Finally the absence of blinding can distort the comparability of the groups regarding study withdrawal or patients' adherence, since patients participating in a non-inferiority trial may prefer to receive the simpler therapy. Among the studies observed, significantly more patients discontinued the ALIZE study medication in the control arm for personal reasons, as compared with the simpler, once-a-day experimental group (11% versus 2%, $P < 0.0004$). This may influence outcome, particularly in an ITT analysis, where withdrawals are considered as failure. Another example comes from the results of the 934 study, where adherence to treatment differed significantly between groups. The conclusion about superior efficacy of the experimental arm in the 934-study may be in part the consequence of greater exposure to the experimental drug. On the other hand, blinding can stand in the way of an optimal drug dispensation in non-inferiority and equivalence trials, in particular if the aim is to simplify antiretroviral therapy. For example, if the purpose is to offer simpler dosage or fewer pills as compared to standard therapy, blinding may require similar regimens in both arms so that any advantages of simplification would be eliminated.

Exclusion of patients after they have been randomized sacrificed the validity of "on-treatment" analysis because it may cause major bias regarding group comparability. For this reason, intention-to-treat analyses has been recognized as the most appropriate and conservative strategy to analyse data of double-blinded trials. However, in case of non-inferiority and equivalence trials, it is well known that this method lacks of robustness since not conservative. For this reason, the study interpretation should also be complemented by "on-treatment analysis"[1,8,9]. If there are discrepancies in the results regarding equivalence or non-inferiority, this should be reported and acknowledged. The CNAAB3005 illustrated how apparent equivalence can be the consequence of a dilutional effect of comparing 2 treatments in the ITT (527 patients) when only 54% of the patients where on-treatment. The same could apply to the ESS40013 study. The use of an "overall" log-rank testing superiority within the 3 arms in the NEFA study may also have blurred the lower efficacy of one study arm, as demonstrated by the "head-to-head" comparison between abacavir and efavirenz.

Like in superiority trials, the choice of the primary outcome is also critical in non-inferiority trials. The BMS-045 illustrated how statistical non-inferiority for viral log difference can be compatible with up to 20.4% of additional virologic failure in the experimental arm, a percentage much larger than non-inferiority margins usually selected for this outcome in this setting.

Finally, the majority of the studies concluded that the effect of at least one experimental arm, based on their prespecified margin, was similar to the control. However, only half of these studies actually demonstrated non-inferiority. Prespecifying the non-inferiority or equivalence margin is necessary but not sufficient to guaranty methodologic quality and appropriate conclusion. We confirmed that AIDS trialists had low adherence to non-inferiority and equivalence methodological standards, as it is the case in other fields[28]. An antiretroviral drug may not prove non-inferiority in term of efficacy but nonetheless be a good alternative because the observed difference is small and the new drug demonstrates better tolerance. This interpretation should, however, be left to the reader. To allow a risk-benefit assessment to be made, the report has a particular obligation to be as clear as possible, using standard statistical vocabulary for non-inferiority and equivalence trials, in compliance with the CONSORT statement.

## Conclusion

Conclusions about non-inferiority should be drawn on the basis of an appropriate confidence interval using a predefined criterion for non-inferiority, shown in both OT and ITT in compliance with the non-inferiority and equivalence extension of the CONSORT statement[1]. We describe how failure to do so will lead to erroneous conclusions. A claim of non-inferiority with a non-inferiority chi-square D-value above 5% is as incorrect as a claim of superiority with traditional null hypothesis testing P-value above 5%. Although the 95% confidence approach is sufficient to reject the null hypothesis, the non-inferiority chi-square provides additional information about the actual degree of significance. Of note, the revised CONSORT statement for superiority trials, item 12a[32] recommends the report of the actual P-values for statistical significance rather than the imprecise threshold "$P < 0.05$". The additional use of the continuity-corrected non-inferiority chi-square may contribute to avoid misleading interpretation by non-statisticians, for whom significance testing may have a higher impact than confidence intervals. The clinical relevance of the primary outcome on which non-inferiority rely should also be assessed. Reviewers and Editors need to reinforce their standards for acceptance of non-inferiority and equivalence randomized controlled trial. Finally, the importance of critical appraisal has implications for both curricular planning in

schools and colleges of medicine, as well as for continuing education programs.

## Competing interests

JJP received research or travel grants from Boehringer Ingelheim, GlaxoSmithKline, Abbott Pharmaceutical, Roche Pharma, GileadSciences. RV received research or travel grants from Bristol-Myers-Squibb, Merck, Boehringer Ingelheim, GlaxoSmithKline, Abbott Pharmaceutical, Roche Pharma, Pfizer and GileadSciences. VM declares that she has no competing interests.

## Authors' contributions

No persons apart from the authors contributed to this paper. JJP had the original idea for the paper. JJP and VM performed the literature search, conducted quality assessment and data extraction and performed statistical analysis. The paper was drafted by JJP and critically appraised for intellectual content by RV and VM, who were also involved in interpretation of the data. All authors read and approved the final manuscript. The guarantor of this paper is JJP.

## Acknowledgements

We would like to thank David Sackett for his useful comments on an earlier version of this manuscript, Jean-Michel Molina and Joel Gallant for helpful discussions on their work. There is no funding for this study.

## References

1. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ: **Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement.** *Jama* 2006, **295:**1152-1160.
2. Temple R, Ellenberg SS: **Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues.** *Ann Intern Med* 2000, **133:**455-463.
3. Djulbegovic B, Clarke M: **Scientific and ethical issues in equivalence trials.** *Jama* 2001, **285:**1206-1208.
4. Staszewski S, Morales-Ramirez J, Tashima KT, Rachlis A, Skiest D, Stanford J, Stryker R, Johnson P, Labriola DF, Farina D, Manion DJ, Ruiz NM: **Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. Study 006 Team.** *N Engl J Med* 1999, **341:**1865-1873.
5. Walmsley S, Bernstein B, King M, Arribas J, Beall G, Ruane P, Johnson M, Johnson D, Lalonde R, Japour A, Brun S, Sun E: **Lopinavir-ritonavir versus nelfinavir for the initial treatment of HIV infection.** *N Engl J Med* 2002, **346:**2039-2046.
6. Parienti JJ: **Tenofovir, equivalence, and noninferiority.** *Jama* 2004, **292:**1951; author reply 1951-2.
7. Kirshner B: **Methodological standards for assessing therapeutic equivalence.** *J Clin Epidemiol* 1991, **44:**839-849.
8. Jones B, Jarvis P, Lewis JA, Ebbutt AF: **Trials to assess equivalence: the importance of rigorous methods.** *Bmj* 1996, **313:**36-39.
9. McAlister FA, Sackett DL: **Active-control equivalence trials and antihypertensive agents.** *Am J Med* 2001, **111:**553-558.
10. Staszewski S, Keiser P, Montaner J, Raffi F, Gathe J, Brotas V, Hicks C, Hammer SM, Cooper D, Johnson M, Tortell S, Cutrell A, Thorborn D, Isaacs R, Hetherington S, Steel H, Spreen W: **Abacavir-lamivudine-zidovudine vs indinavir-lamivudine-zidovudine in antiretroviral-naive HIV-infected adults: A randomized equivalence trial.** *Jama* 2001, **285:**1155-1163.
11. Martinez E, Arnaiz JA, Podzamczer D, Dalmau D, Ribera E, Domingo P, Knobel H, Riera M, Pedrol E, Force L, Llibre JM, Segura F, Richart C, Cortes C, Javaloyas M, Aranda M, Cruceta A, de Lazzari E, Gatell JM: **Substitution of nevirapine, efavirenz, or abacavir for protease inhibitors in patients with human immunodeficiency virus infection.** *N Engl J Med* 2003, **349:**1036-1046.
12. Arnaiz JA, Mallolas J, Podzamczer D, Gerstoft J, Lundgren JD, Cahn P, Fatkenheuer G, D'Arminio-Monforte A, Casiro A, Reiss P, Burger DM, Stek M, Gatell JM: **Continued indinavir versus switching to indinavir/ritonavir in HIV-infected patients with suppressed viral load.** *Aids* 2003, **17:**831-840.
13. van Leth F, Phanuphak P, Ruxrungtham K, Baraldi E, Miller S, Gazzard B, Cahn P, Lalloo UG, van der Westhuizen IP, Malan DR, Johnson MA, Santos BR, Mulcahy F, Wood R, Levi GC, Reboredo G, Squires K, Cassetti I, Petit D, Raffi F, Katlama C, Murphy RL, Horban A, Dam JP, Hassink E, van Leeuwen R, Robinson P, Wit FW, Lange JM: **Comparison of first-line antiretroviral therapy with regimens including nevirapine, efavirenz, or both drugs, plus stavudine and lamivudine: a randomised open-label trial, the 2NN Study.** *Lancet* 2004, **363:**1253-1263.
14. Gallant JE, Staszewski S, Pozniak AL, DeJesus E, Suleiman JM, Miller MD, Coakley DF, Lu B, Toole JJ, Cheng AK: **Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial.** *Jama* 2004, **292:**191-201.
15. Gathe JCJ, Ive P, Wood R, Schurmann D, Bellos NC, DeJesus E, Gladysz A, Garris C, Yeo J: **SOLO: 48-week efficacy and safety comparison of once-daily fosamprenavir /ritonavir versus twice-daily nelfinavir in naive HIV-1-infected patients.** *Aids* 2004, **18:**1529-1537.
16. Benson CA, van der Horst C, Lamarca A, Haas DW, McDonald CK, Steinhart CR, Rublein J, Quinn JB, Mondou E, Rousseau F: **A randomized study of emtricitabine and lamivudine in stably suppressed patients with HIV.** *Aids* 2004, **18:**2269-2276.
17. DeJesus E, McCarty D, Farthing CF, Shortino DD, Grinsztejn B, Thomas DA, Schrader SR, Castillo SA, Sension MG, Gough K, Madison SJ: **Once-daily versus twice-daily lamivudine, in combination with zidovudine and efavirenz, for the treatment of antiretroviral-naive adults with HIV infection: a randomized equivalence trial.** *Clin Infect Dis* 2004, **39:**411-418.
18. Molina JM, Journot V, Morand-Joubert L, Yeni P, Rozenbaum W, Rancinan C, Fournier S, Morlat P, Palmer P, Dupont B, Goujard C, Dellamonica P, Collin F, Poizot-Martin I, Chene G: **Simplification therapy with once-daily emtricitabine, didanosine, and efavirenz in HIV-1-infected adults with viral suppression receiving a protease inhibitor-based regimen: a randomized trial.** *J Infect Dis* 2005, **191:**830-839.
19. DeJesus E, Herrera G, Teofilo E, Gerstoft J, Buendia CB, Brand JD, Brothers CH, Hernandez J, Castillo SA, Bonny T, Lanier ER, Scott TR: **Abacavir versus zidovudine combined with lamivudine and efavirenz, for the treatment of antiretroviral-naive HIV-infected adults.** *Clin Infect Dis* 2004, **39:**1038-1046.
20. Squires K, Lazzarin A, Gatell JM, Powderly WG, Pokrovskiy V, Delfraissy JF, Jemsek J, Rivero A, Rozenbaum W, Schrader S, Sension M, Vibhagool A, Thiry A, Giordano M: **Comparison of once-daily atazanavir with efavirenz, each in combination with fixed-dose zidovudine and lamivudine, as initial therapy for patients infected with HIV.** *J Acquir Immune Defic Syndr* 2004, **36:**1011-1019.
21. Markowitz M, Hill-Zabala C, Lang J, DeJesus E, Liao Q, Lanier ER, Davis EA, Shaefer M: **Induction with abacavir/lamivudine/zidovudine plus efavirenz for 48 weeks followed by 48-week maintenance with abacavir/lamivudine/zidovudine alone in antiretroviral-naive HIV-1-infected patients.** *J Acquir Immune Defic Syndr* 2005, **39:**257-264.
22. Sosa N, Hill-Zabala C, Dejesus E, Herrera G, Florance A, Watson M, Vavro C, Shaefer M: **Abacavir and lamivudine fixed-dose combination tablet once daily compared with abacavir and lamivudine twice daily in HIV-infected patients over 48 weeks (ESS30008, SEAL).** *J Acquir Immune Defic Syndr* 2005, **40:**422-427.
23. Johnson M, Grinsztejn B, Rodriguez C, Coco J, DeJesus E, Lazzarin A, Lichtenstein K, Rightmire A, Sankoh S, Wilber R: **Atazanavir plus ritonavir or saquinavir, and lopinavir/ritonavir in patients experiencing multiple virological failures.** *Aids* 2005, **19:**685-694.
24. Moyle GJ, DeJesus E, Cahn P, Castillo SA, Zhao H, Gordon DN, Craig C, Scott TR: **Abacavir once or twice daily combined with once-daily lamivudine and efavirenz for the treatment of antiretroviral-naive HIV-infected adults: results of the Ziagen Once**

**Daily in Antiretroviral Combination Study.** *J Acquir Immune Defic Syndr* 2005, **38:**417-425.
25. **Telzir European European Public Assessment Report. Scientific discussion http://www.emea.eu.int/humandocs/Humans/EPAR/telzir/telzir.htm.** .
26. Bonjoch A, Paredes R, Galvez J, Miralles C, Videla S, Martinez E, Miranda J, Munoz-Moreno JA, De la Torre J, Prieto A, Vilades C, Clotet B: **Antiretroviral treatment simplification with 3 NRTIs or 2 NRTIs plus nevirapine in HIV-1-infected patients treated with successful first-line HAART.** *J Acquir Immune Defic Syndr* 2005, **39:**313-316.
27. Gallant JE, DeJesus E, Arribas JR, Pozniak AL, Gazzard B, Campo RE, Lu B, McColl D, Chuck S, Enejosa J, Toole JJ, Cheng AK: **Tenofovir DF, emtricitabine, and efavirenz vs. zidovudine, lamivudine, and efavirenz for HIV.** *N Engl J Med* 2006, **354:**251-260.
28. Le Henanff A, Giraudeau B, Baron G, Ravaud P: **Quality of reporting of noninferiority and equivalence randomized trials.** *Jama* 2006, **295:**1147-1151.
29. Dunnett CW, Gent M: **Significance testing to establish equivalence between treatments, with special reference to data in the form of 2X2 tables.** *Biometrics* 1977, **33:**593-602.
30. Dunnett CW, Gent M: **An alternative to the use of two-sided tests in clinical trials.** *Stat Med* 1996, **15:**1729-1738.
31. Snapinn SM: **Noninferiority trials.** *Curr Control Trials Cardiovasc med* 2000, **1:**19-21.
32. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T: **The revised CONSORT statement for reporting randomized trials: explanation and elaboration.** *Ann Intern Med* 2001, **134:**663-694.

## Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/6/46/prepub