

Reliability evaluation of four different assays for therapeutic drug monitoring of infliximab levels

Irene Pérez, Lidia Fernández, Silvia Sánchez-Ramón, Cristina Alba, Ana Zatarain, Mercedes Cañas, Olga N. López, David Olivares, Enrique Rey and Carlos Taxonera 

Ther Adv Gastroenterol

2018, Vol. 11: 1–10

DOI: 10.1177/
1756284818783613

© The Author(s), 2018.
Reprints and permissions:
[http://www.sagepub.co.uk/
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

Abstract

Background: The aim of this study was to evaluate reliability of four different assays for measuring infliximab trough levels and antibodies to infliximab (ATI).

Methods: In this non-interventional, cross-sectional study including IBD patients, infliximab levels and ATI were measured using four different assays: Lisa-Tracker, Promonitor, Q-Inflix and Sanquin. Reliability and agreement for infliximab levels was assessed using the intraclass correlation coefficient (ICC) and Bland–Altman plots. Qualitative agreement for infliximab (based on a pre-established target window of trough levels between 3 µg/ml and 7 µg/ml) and for ATI were estimated by Cohen’s kappa.

Results: Serum samples of 84 IBD patients were evaluated for infliximab using the four assays. Reliability was ‘substantial’ between Lisa-Tracker *versus* Promonitor and ‘almost perfect’ between the remaining assay pairs, with ICCs [95% confidence interval (CI)] ranging from 0.93 (0.70–0.97) for Lisa-Tracker *versus* Promonitor to 0.97 (0.95–0.98) for Q-Inflix *versus* Sanquin. Bland–Altman plots showed significant bias between assays except Promonitor *versus* Q-Inflix, which had excellent agreement. The greatest differences in mean infliximab were found between Promonitor *versus* Lisa-Tracker (–0.91 µg/ml) and Lisa-Tracker *versus* Q-Inflix (0.69 µg/ml). Qualitative agreement for infliximab was ‘almost perfect’ for Promonitor *versus* Q-Inflix (kappa 0.84) and Q-Inflix *versus* Sanquin (kappa 0.81), and ‘substantial’ for the remaining pairs. More than 10% of patients who had infliximab levels within the target interval by Lisa-Tracker had suboptimal concentrations (<3 µg/ml) with Promonitor and Q-Inflix. Furthermore, 11% of patients within the target interval by Q-Inflix had supra-optimal levels (>7 µg/ml) by Lisa-Tracker. In the remaining paired comparisons, fewer than 5% of patients were placed in different subgroups. Qualitative agreement for ATI fluctuated between ‘moderate’ and ‘almost perfect’.

Conclusions: All four assays seem suitable for therapeutic drug monitoring of infliximab. Promonitor and Q-Inflix had the best agreement, making those assays fully interchangeable. Systematic biases between Lisa-Tracker with Promonitor and Q-Inflix suggest that these assays should not be interchanged during the follow up of an individual patient.

Keywords: antibodies to infliximab, inflammatory bowel disease, infliximab, intraclass correlation coefficient, reliability, therapeutic drug monitoring, trough levels

Received: 23 January 2018; revised manuscript accepted: 24 May 2018.

Introduction

Infliximab (IFX), a chimeric mouse/human monoclonal antibody against tumour necrosis factor- α , is highly effective in inducing and maintaining

remission in patients with Crohn’s disease and ulcerative colitis.^{1,2} During IFX maintenance therapy, a substantial fraction of patients have secondary loss of response.^{3–5} Loss of response is overcome

Correspondence to:
Carlos Taxonera
Inflammatory Bowel
Disease Unit, Department
of Gastroenterology,
Hospital Clínico San
Carlos, c/ Profesor
Martín Lagos s/n, Madrid,
28040, Spain
Instituto de Investigación
del Hospital Clínico San
Carlos (IdISSC), Madrid,
Spain
carlos.taxonera@salud.madrid.org

Irene Pérez
Inflammatory Bowel
Disease Unit, Department
of Gastroenterology,
Hospital Clínico San
Carlos, Madrid, Spain
Department of
Gastroenterology, Hospital
La Paz, Madrid, Spain

Lidia Fernández
Silvia Sánchez-Ramón
Department of Clinical
Immunology, Hospital
Clínico San Carlos, Madrid,
Spain
Instituto de Investigación
del Hospital Clínico San
Carlos (IdISSC), Madrid,
Spain

Cristina Alba
Ana Zatarain
Mercedes Cañas
Olga N. López
Inflammatory Bowel
Disease Unit, Department
of Gastroenterology,
Hospital Clínico San
Carlos, Madrid, Spain

David Olivares
Enrique Rey
Inflammatory Bowel
Disease Unit, Department
of Gastroenterology,
Hospital Clínico San
Carlos, Madrid, Spain
Instituto de Investigación
del Hospital Clínico San
Carlos (IdISSC), Madrid,
Spain

by IFX dose escalation in most cases, although a subset of patients had IFX failure despite dose escalation. Possible reasons for IFX failure include: (1) non-immune related factors such as high body mass index or high disease burden, which increase drug clearance;^{6,7} (2) development of antidrug antibodies, which form immune complexes and reduce exposure of the drug and increase its clearance;^{7,8} and (3) tumour necrosis factor-independent disease mechanisms.⁹

Several studies show that serum IFX trough levels are correlated with clinical response, clinical remission and mucosal healing in patients with inflammatory bowel disease (IBD).^{10–12} Recently, an optimal therapeutic window of 3–7 µg/ml has been reported.¹³ Inadequate serum drug concentrations and antidrug antibodies are associated with poor clinical outcomes.^{14–16} Therapeutic drug monitoring may help to identify mechanisms for loss of response and to guide selection of optimal intervention in individual patients. The use of therapeutic monitoring to guide treatment decisions in patients with loss of response to IFX has been shown to be cost-effective compared with empiric dose escalation.^{17,18}

Different analytical techniques are currently used to measure IFX levels and antibodies to IFX (ATI), including enzyme-linked immunosorbent assay (ELISA), radio-immunoassay (RIA), functional cell-based reporter gene assay and homogeneous mobility shift assays. However, there is no gold standard technique available and assay heterogeneity may constitute an important bias for clinical application of therapeutic drug monitoring, hampering comparison of results from different studies.¹⁹ Although several studies reported systematic differences in exact drug concentrations, even between assays that share the same technology, the different analytical techniques for quantification of IFX levels and ATI result in similar classifications and therapeutic interventions in most patients with IFX failure, leading to comparable clinical outcomes.²⁰

The aim of this study was to assess the reliability of four different assays for measuring IFX levels and ATI in patients with IBD under maintenance treatment. In this study we compared the assays developed by Theradiag (Lisa-Tracker Infliximab and Lisa-Tracker Anti-Infliximab), Progenika Biopharma (Promonitor-IFX V.2 and Promonitor-Anti-IFX V.2), Matriks Biotek (Q-Inflix-ELISA and Q-ATI ELISA) and Sanquin Diagnostics (Sanquin IFX ELISA and RIA).

Material and methods

Study population and sample collection

This was a single-centre, non-interventional, cross-sectional study. The study was approved by the ethics committee of the centre (CI.12/148). Informed consent was obtained from each participant. The study population comprised all consecutive IBD patients who were primary responders to IFX induction doses and who received at least three IFX maintenance doses. Demographic and clinical characteristics, concomitant treatment, as well as dosing of IFX (milligrams per kilogram and interval between infusions) were recorded. Samples were taken immediately before the next intravenous infusion of IFX. Venous blood samples were collected into BD Vacutainer serum tubes, and processed immediately or within 4 h after extraction. Serum samples were obtained after centrifugation at 2330 g for 10 min at room temperature, and finally four separated aliquots were frozen at –40°C.

Measurement of infliximab trough levels and antibodies to infliximab

We performed the quantification of IFX levels and ATI with the three commercially available ELISA kits in Spain: Lisa-Tracker Infliximab and Lisa-Tracker Anti-Infliximab (Theradiag, France), Promonitor-IFX V.2 and Promonitor-ANTI-IFX V.2 (Progenika Biopharma, Spain) and Q-Inflix-ELISA and Q-ATI ELISA (Matriks Biotek, Turkey). We also sent aliquots to Sanquin Diagnostics (Netherlands) for measurement of IFX and ATI with in-house assays that have been used for several years in clinical practice. All kits were used by a qualified person in accordance with the manufacturer's instructions. Drug-level analyses for all kits were carried out using a capture or sandwich ELISA, and ATI were measured using bridging ELISA (Lisa-Tracker and Promonitor), capture ELISA (Q-Inflix) and RIA (Sanquin), as previously described.^{21–24} The lower limit of quantification for IFX concentrations was 0.1, 0.03, 0.02 and 0.002 µg/ml for Lisa-Tracker, Promonitor, Q-Inflix and Sanquin, respectively. The upper limit of quantification for IFX concentrations was 16, 14.4, 20 and 47 µg/ml for Lisa-Tracker, Promonitor, Q-Inflix and Sanquin assays, respectively. The lower and upper limits of quantification of ATI assays were: Lisa-Tracker (10–200 µg/ml), Promonitor (2–1440 AU/ml), Q-Inflix (30–500 µg/ml) and

Sanquin (in-house method, concentrations >12 antibody units/ml were classified as positive).

Statistical analysis

Results for IFX levels are summarized as median and interquartile range (IQR), and qualitative data are shown as number and percentages. To quantify the reliability for IFX levels from all four assays, the intraclass correlation coefficient (ICC) was calculated using the two-way mixed-effects single-rater model (absolute agreement) whereby a value of 1 represents complete agreement. Pairwise ICC comparisons were also performed. We categorized degrees of reliability of ICC based on the lowest 95% confidence interval (CI) as: 0.21–0.4 ‘low’; 0.41–0.6 ‘moderate’; ‘0.61–0.8’ substantial; and >0.81 ‘almost perfect’.²⁵

Bland–Altman plots in which the difference between two measurements is plotted on the y-axis, and the average of two measurements on the x-axis were drawn for each pair of methods. This plot allows comparison of agreement of two assay methods. An ideal agreement between two assay methods is represented by a flat line in the Bland–Altman plot. Mean difference and its 95% CI were calculated. Limits of agreement were defined as mean difference \pm two standard deviations. Ninety-five percent of differences are expected to lie between limits of agreement. The qualitative agreement between IFX levels was determined using Cohen’s weighted kappa. Kappa statistics were calculated based on a pre-established target window of IFX trough levels between 3 μ g/ml and 7 μ g/ml.¹³ ATI were classified as detectable/undetectable. Fleiss kappa determined qualitative agreement between the four methods. Qualitative agreement between pairs of methods was determined using Cohen’s kappa based on the lowest 95% CI and categorized according to the criteria of Landis and Koch.²⁶ *p* values <0.05 were considered significant. Statistical analysis was performed using SPSS 21.0 and STATA 12.0.

Results

Study population

Between May and July 2016, serum samples were collected from 86 IBD patients. IFX concentrations in samples from two patients were above the upper limit of quantification in the Lisa-Tracker and Promonitor assays and also gave a high signal

with the Q-Inflix and Sanquin assays. These samples were excluded from any further calculations. Baseline characteristics of the 84 evaluable patients are summarized in Table 1.

Infliximab trough levels

Quantitative analysis. IFX median (IQR) trough levels were 4.33 μ g/ml (1.00–4.80) with Lisa-Tracker, 2.4 μ g/ml (1.75–6.52) with Promonitor, 2.61 μ g/ml (0.86–4.81) with Q-Inflix and 3.40 μ g/ml (1.50–5.60) with Sanquin. IFX levels determined by Lisa-Tracker were frequently higher than the corresponding Sanquin levels, whereas Promonitor and Q-Inflix show levels frequently lower than the corresponding Sanquin determinations. Linear correlation was excellent between all assay pairs (Pearson’s *r* correlation coefficient: 0.95 Promonitor *versus* Sanquin and Q-Inflix *versus* Sanquin; 0.94 Lisa-Tracker *versus* Promonitor; 0.92 Promonitor *versus* Q-Inflix; and 0.91 Lisa-Tracker *versus* Sanquin; *p* < 0.001 for all comparisons).

The overall inter-assay reliability of the four different assays for IFX levels was studied using the ICC. The ICC for absolute agreement was ‘almost perfect’ (0.97, 95% CI 0.96–0.98) between the four assays. To study possible inter-assay differences, a pairwise comparison was performed. Reliability was ‘substantial’ between Lisa-Tracker and Promonitor and ‘almost perfect’ between the remaining assay pairs, with ICCs ranging between 0.93 (95% CI 0.70–0.97) and 0.97 (95% CI 0.95–0.98) (Table 2).

Bland–Altman plots for each assay pair and values for mean differences and limits of agreement are shown in Figure 1. For all assay pairs except one, the 95% CI of the mean differences of infliximab levels did not include zero. The mean difference between Promonitor and Q-Inflix was not significantly different from zero (–0.15 μ g/ml, 95% CI: –0.43–0.14; *p* = 0.306). The Bland–Altman plots denoted systematic biases with higher IFX levels in Lisa-Tracker than in the other three assays, and higher IFX levels in Sanquin than in Promonitor and Q-Inflix. The highest mean differences were found between Promonitor *versus* Lisa-Tracker (–0.91 μ g/ml, 95% CI –1.18–0.63; *p* < 0.001) and Lisa-Tracker *versus* Q-Inflix (0.69 μ g/ml, 95% CI 0.40–0.98; *p* < 0.001), showing that the concentrations measured by Promonitor and Q-Inflix were consistently lower

Table 1. Baseline characteristics of patients.

Patients, <i>n</i>	84
Males, <i>n</i> (%)	52 (61.9)
Age, year, median (IQR)	45.5 (37–55)
Crohn's disease, <i>n</i> (%)	58 (69)
Ulcerative colitis, <i>n</i> (%)	26 (31)
CD: location, <i>n</i> (%)*	
L1 (terminal ileum)	23 (39.7) **
L2 (colon)	11 (18.9)
L3 (ileocolon)	22 (37.9) ***
Perianal	28 (48.3)
CD: behaviour, <i>n</i> (% of CD)*	
B1 (non-stricturing, non-penetrating)	27 (48.2)
B2 (stricturing)	10 (17.9)
B3 (penetrating)	19 (33.9)
UC: extension, <i>n</i> (% of UC)	
E1 Proctitis	4 (15.4)
E2 Left-sided	11 (42.3)
E3 Extensive	11 (42.3)
Disease duration, year, median (IQR)	11.5 (7–20)
Smoker status	
No	48 (57.1)
Yes	36 (42.9)
Time since first IFX infusion, months, median (IQR)	36.0 (24.5–59.5)
Concomitant immunosuppressants, <i>n</i> (%)	39 (46.4)
C-reactive protein <0.5 mg/dl, <i>n</i> (%)	48 (69)
Serum albumin (g/dl), median (IQR)	4.2 (4.1–4.5)
CD, Crohn's disease; IFX, infliximab; IQR, interquartile range; UC: ulcerative colitis. *Two patients not characterized. Disease could be recorded in more than one location. **Three patients with upper GI involvement. ***One patient with upper GI involvement.	

than those measured by Lisa-Tracker. When the average of two measurements increased above 7 µg/ml, the difference in measurements between assays also increased for all pairwise comparisons. In the concentration range of infliximab between

0 µg/ml and 7 µg/ml, differences in most assays were small, with the exception of the comparison between Lisa-Tracker with Promonitor and Q-Inflixi assays.

Table 2. Intraclass correlation coefficients (95% confidence interval) for quantitative agreement of infliximab trough levels: comparison between assays pairs.

Assay	Promonitor	Q-Inflix	Sanquin
Lisa-Tracker	0.93 (0.70–0.97)	0.95 (0.87–0.97)	0.95 (0.92–0.97)
Promonitor		0.94 (0.91–0.96)	0.95 (0.89–0.97)
Q-Inflix			0.97 (0.95–0.98)

Qualitative analysis. The results for each assay were stratified according to the pre-established target window of IFX trough levels between 3 µg/ml and 7 µg/ml. IFX <3 µg/ml and >7 µg/ml were classified as suboptimal or supraoptimal, respectively.¹³ An ‘almost perfect’ agreement was found between Promonitor and Q-Inflix ($k = 0.840$) and Lisa-Tracker and Q-Inflix ($k = 0.811$). Comparisons between the remaining assay pairs showed a ‘substantial’ agreement (Table 3). Lisa-Tracker assay yielded the highest IFX concentrations: 11% and 14% of the patients who were within the target interval of 3–7 µg/ml when evaluated by Lisa-Tracker had suboptimal drug concentrations (<3 µg/ml) when measured with Promonitor and Q-Inflix, respectively. Furthermore, 11% of the patients considered within the target interval by Q-Inflix had supraoptimal levels (>7 µg/ml) when measured by Lisa-Tracker. In the remaining paired comparisons, fewer than 5% of patients were placed in different subgroups. We carried out a sensitivity analysis considering a minimum IFX trough level of ≥ 5 µg/ml as the therapeutic threshold.²⁷ Again, Lisa-Tracker yielded the highest IFX concentrations: 13% and 14% of the patients with an IFX trough level of ≥ 5 µg/ml when evaluated by Lisa-Tracker had suboptimal drug concentrations (<5 µg/ml) when measured with Promonitor and Q-Inflix. In the remaining paired comparisons, less than 5% of patients were discordant for the threshold of ≥ 5 µg/ml.

Antibodies to infliximab

For comparisons between ATI assays, only 36 serum samples tested with the four assays were considered. Four serum samples had detectable ATIs and 30 serum samples were ATI-negative in all four assays; thus, only two samples had a discordant inter-assay classification [Fleiss kappa 0.870 ($p < 0.001$)]. The two discordant samples were ATI-positive with Lisa-Tracker and Sanquin

and ATI-negative with Promonitor and Q-Inflix. Pairwise comparisons by Cohen’s kappa showed ‘almost perfect’ agreement between the pairs of assays Lisa-Tracker and Sanquin, ‘substantial’ agreement for assay Lisa-Tracker and Q-Inflix and Q-Inflix and Sanquin, and ‘moderate’ agreement between the remaining pairs (Table 4). There were detectable levels of infliximab in the 30 samples with undetectable ATIs with the four assays.

Discussion

Several methods can be used to measure IFX levels and ATI, including ELISA, RIA, functional cell-based reporter gene assay and homogeneous mobility shift assays, although ELISA-based assays are most commonly used given their cost and practicality. The variability in results between methodologies makes it difficult to compare data from different studies. At present there is unfortunately no defined gold standard for quantification of IFX levels and ATI.¹⁹ Therefore, comparing assays is critical for understanding and interpreting data of different clinical studies and before they are used for clinical decision-making based on a predefined therapeutic algorithm.²¹

In this study we evaluated IBD patients receiving IFX maintenance treatment with four ELISA-based assays for IFX trough levels, and with three ELISA-based assays and one RIA-based assay for detection of ATI. All these assays are frequently used for therapeutic drug monitoring and therefore inter-assay reliability analyses are desirable. In the case of IFX trough levels, the overall comparison between all four ELISAs showed an ‘almost perfect’ ICC for absolute agreement, suggesting that the four assays could be used for therapeutic drug monitoring of infliximab. Pairwise comparison using a strict categorization based on lower 95% CI confirmed an ‘almost perfect’ reliability between all but one assay pair. Reliability

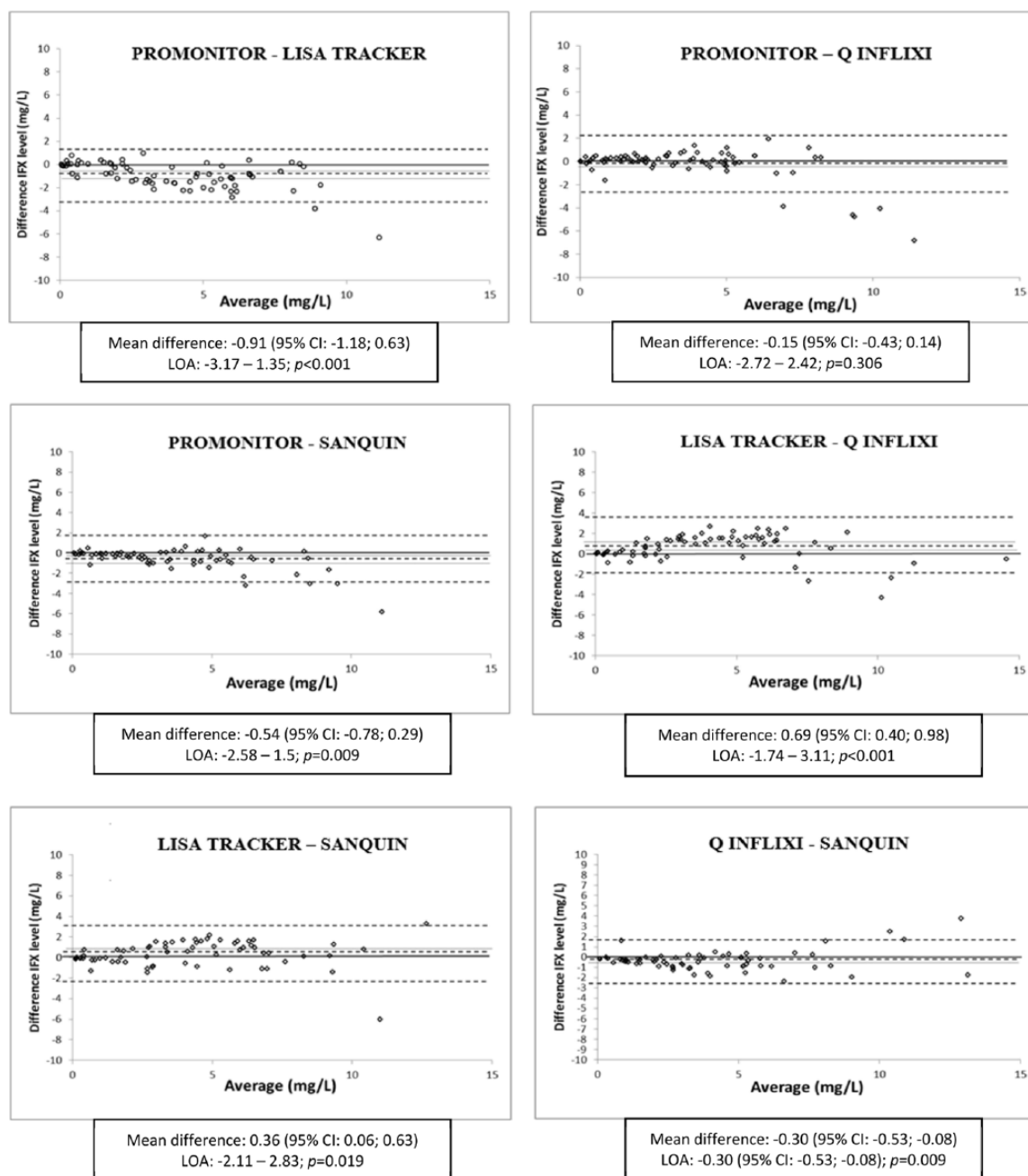


Figure 1. Bland–Altman plots of infliximab levels between assay pairs. The differences between each assay are plotted on the y-axis and their averages are plotted on the x-axis. The central dotted lines represents the mean difference and the closer dotted line are the upper and lower limits of the 95% confidence interval (CI). The more external dotted lines are the upper and lower limits of agreement (LOA).

between Lisa-Tracker and Promonitor assays was only ‘substantial’.

Further analysis evidenced some relevant differences between assays. Bland–Altman plots show a greater dispersion in all paired comparisons when

IFX concentrations were over 7 $\mu\text{g/ml}$. The differences increase with increasing IFX concentrations, as described previously in other comparison studies.^{21,28} The Bland–Altman plot comparing Promonitor and Q-Inflixi showed excellent agreement, making those assays fully interchangeable.

Table 3. Qualitative agreement (Cohen's weighted kappa) between categorized infliximab trough levels^{**}: comparison between assays pairs.

Assays	Observed agreement	Expected agreement	Cohen's kappa
Lisa-Tracker <i>versus</i> Promonitor	88.0%	62.8%	0.677*
Lisa-Tracker <i>versus</i> Q-Inflixix	87.3%	60.6%	0.678*
Lisa-Tracker <i>versus</i> Sanquin	87.5%	61.9%	0.672*
Promonitor <i>versus</i> Q-Inflixix	94.6%	66.5%	0.840*
Promonitor <i>versus</i> Sanquin	93.0%	65.5%	0.796*
Q-Inflixix <i>versus</i> Sanquin	93.0%	62.6%	0.812*

* $p < 0.001$.
^{**}Infliximab target interval: 3–7 µg/ml; infliximab concentrations <3 µg/ml and >7 µg/ml were classified as suboptimal or supraoptimal, respectively.

Table 4. Qualitative agreement (Cohen's kappa) for antibodies to infliximab: comparison between assay pairs.

Assays	Positive agreement	Negative agreement	Overall agreement	Cohen's kappa (95% CI)
Lisa-Tracker <i>versus</i> Promonitor	14.3%	79.6%	93.9%	0.788 (0.558–1.000)*
Lisa-Tracker <i>versus</i> Q-Inflixix	87.7%	9.9%	97.6%	0.875 (0.705–1.000)*
Lisa-Tracker <i>versus</i> Sanquin	91.2%	8.8%	100%	1.000*
Promonitor <i>versus</i> Q-Inflixix	12.2%	81.6%	93.8%	0.764 (0.507–1.000)*
Promonitor <i>versus</i> Sanquin	83.3%	11.1%	94.4%	0.769 (0.467–1.000)*
Q-Inflixix <i>versus</i> Sanquin	91.5%	7.0%	98.5%	0.902 (0.711–1.000)*

* $p < 0.001$.

Conversely, due to the low agreement, our results indicate that Lisa-Tracker should not be interchanged with Promonitor and Q-Inflixix during the follow up of individual patients. For infliximab concentrations between 0 µg/ml and 7 µg/ml, differences were small between the remaining pairs of assays.

In the qualitative analyses, the results for each assay were stratified according to a recommended therapeutic window for patients receiving IFX maintenance treatment. The TAXIT trial demonstrated that dose escalation in IBD patients with

suboptimal drug concentrations (<3 µg/ml) results in better disease control, while dose reduction was successful in IBD patients with supraoptimal drug concentrations (>7 µg/ml), which resulted in a decrease in costs and also fewer adverse events.¹³ There was an 'almost perfect' agreement between Q-Inflixix with Promonitor and Sanquin assays, and a 'substantial' agreement between the remaining assay pairs. However, 11% and 14% of the patients with suboptimal drug concentrations when measured by Promonitor and Q-Inflixix, were within the target interval of 3–7 µg/ml when evaluated by Lisa-Tracker. In the

remaining paired comparison, fewer than 5% of patients were placed in different subgroups. Recently, the American Gastroenterological Association Institute Technical Review proposed an infliximab therapeutic threshold of ≥ 5 $\mu\text{g/ml}$ during maintenance therapy.²⁷ In our study, the sensitivity analysis considering this threshold offered very similar results.

Although the reliability (ICC) between assays was excellent and kappa statistic showed ‘substantial’ and even ‘almost perfect’ qualitative agreement in some pairs, median absolute concentrations differ from assay to assay. Systematic differences in exact drug concentrations have been described previously with different types of assays and even between assays that share the same format and technology.^{20,29} Due to the aforementioned systematic differences and biases in Bland–Altman plots and disagreement in qualitative analysis, we consider that the four different assays are not interchangeable for the purpose of making therapeutic decisions during the follow up of patients. Although the four evaluated assays placed the vast majority of patients in the same categories based on a predefined therapeutic window, a non-negligible percentage of the patients was classified in different categories, with the possible change of management that this implies. The exception was the Promonitor and Q-Inflix ELISA assays, which seem to be fully interchangeable according to our results.

To date, few studies have been done to assess reliability between different assays measuring IFX trough levels. Despite the limitations and the different sensitivities and problems of each assay, the majority of studies of assays report adequate agreement between assays, classifying the majority of patients in a similar way.²⁰ Vande Castele and colleagues compared three different techniques, including Lisa-Tracker and Sanquin assays, also evaluated in our study.²¹ The authors found a good correlation and agreement for IFX levels and ATI measurements between these assays. However, the authors reported that the Lisa-Tracker assay detected IFX in 18% of samples with undetectable levels by the other two assays, and found a correlation of 0.83 where Sanquin assay measured lower levels than Lisa-Tracker, as observed in our study. Two comparison studies performed by Ruiz-Argüello and colleagues³⁰ and Schmitz and colleagues³¹ found that Sanquin measured higher IFX concentrations than Promonitor, as observed

in our study. Several studies that have compared different assays suggested that the same assay should be used in individual patients during follow up.^{19,32,33} Our data support this suggestion.

The ICCs for ATIs could not be evaluated because the assays used different units of measurement. Therefore, ATIs were classified as detectable/undetectable for all analyses. The value of Fleiss kappa for the four assays was almost perfect, but only samples with available results for the four tests were considered. All the assays evaluated in this study for ATIs are drug-sensitive to some extent and so the ability to detect antibodies may be affected by the presence of IFX. The sensitivity of the four assays was comparable, although the Lisa-Tracker and Sanquin assays detected ATIs that were not detected with the other two assays. It has been reported that, since RIA is less sensitive to drug interference than ELISA, it can detect low ATI concentrations in the presence of drug.²¹ ATI results should be interpreted with caution, given the high rate of missing data (particularly when we compare the four assays together) and the very low number of ATI-positive samples, which makes it difficult to draw conclusions.

Our study had other limitations. We have not tested healthy control sera or IFX-spiked samples. Due to the lack of a gold standard technique, the studies performed with spiked samples would provide the true concentration of the drug. In our study, conclusions regarding reliability for IFX levels depend mainly on the results of the pairwise ICCs. We reported degrees of reliability of ICC based on the lowest 95% CI, which reinforces the quality of the evidence.²⁵ Missing data for IFX levels were less than 5% in all four assays.

In conclusion, in a cohort of IBD patients undergoing IFX maintenance therapy, we found a ‘substantial’ to ‘almost perfect’ quantitative and qualitative reliability for IFX levels measured with all four assays, and all of them seem suitable for therapeutic drug monitoring of IFX. Promonitor and Q-Inflix ELISA had the best agreement, making those assays fully interchangeable. However, we found systematic bias between Lisa-Tracker assay with Promonitor and Q-Inflix assays, which could affect interpretation and suggests these assays are not interchangeable for the purposes of making therapeutic decisions during the follow up of an individual patient. International standards should be developed for the use of

different assays that measure IFX levels and antibodies to IFX.

Acknowledgements

The authors would like to thank Dr J. Garrido and C. Fernandez for their assistance in the statistical analysis and Dr G. Morley for writing support and for reviewing the English-language manuscript.

Author contribution

IP and CT contributed equally to the manuscript. IP conceived and designed the study, participated in patient recruitment and data analysis and interpretation, and drafting of the manuscript. LD and SS-R contributed in the study design, participated in data analysis and interpretation and revised the manuscript critically for important intellectual content. CA, AZ, MC, and ONL participated in patient recruitment and data collection. DO participated in data collection and interpretation and performed the statistical analysis. ER revised the manuscript critically for important intellectual content. CT conceived and designed the study, participated in patient recruitment and data analysis and interpretation, and drafting of the manuscript. All authors read and approved the final version of the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest statement

C. Taxonera has served as a speaker, a consultant and advisory member for or has received research funding from MSD, Abbvie, Pfizer, Takeda, Janssen, Ferring, Faes Farma, Shire Pharmaceuticals, Dr. Falk Pharma and Gebro Pharma. The remaining authors declare that they have nothing to disclose.

ORCID iD

Carlos Taxonera  <https://orcid.org/0000-0001-9166-7350>

References

- Hanauer SB, Feagan BG, Lichtenstein GR, *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 2002; 359: 1541–1549.
- Rutgeerts P, Sandborn WJ, Feagan BG, *et al.* Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005; 353: 2462–2476.
- Ford AC, Sandborn WJ, Khan KJ, *et al.* Efficacy of biological therapies in inflammatory bowel disease: systematic review and meta-analysis. *Am J Gastroenterol* 2011; 106: 644–659.
- Gisbert JP and Panés J. Loss of response and requirement of infliximab dose intensification in Crohn's disease: a review. *Am J Gastroenterol* 2009; 104: 760–767.
- Arias MT, Vande Casteele N, Vermeire S, *et al.* A panel to predict long-term outcome of infliximab therapy for patients with ulcerative colitis. *Clin Gastroenterol Hepatol* 2015; 13: 531–538.
- Ordás I, Mould DR, Feagan BG, *et al.* Anti-TNF monoclonal antibodies in inflammatory bowel disease: pharmacokinetics-based dosing paradigms. *Clin Pharmacol Ther* 2012; 91: 635–646.
- Baert F, Noman M, Vermeire S, *et al.* Influence of immunogenicity on the long-term efficacy of infliximab in Crohn's disease. *N Engl J Med* 2003; 348: 601–608.
- Nanda KS, Cheifetz AS and Moss AC. Impact of antibodies to infliximab on clinical outcomes and serum infliximab levels in patients with inflammatory bowel disease (IBD): a meta-analysis. *Am J Gastroenterol* 2013; 108: 40–47.
- Ordás I, Feagan BG and Sandborn WJ. Therapeutic drug monitoring of tumor necrosis factor antagonists in inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2012; 10: 1079–1087.
- Maser EA, Villela R, Silverberg MS, *et al.* Association of trough serum infliximab to clinical outcome after scheduled maintenance treatment for Crohn's disease. *Clin Gastroenterol Hepatol* 2006; 4: 1248–1254.
- Seow CH, Newman A, Irwin SP, *et al.* Trough serum infliximab: a predictive factor of clinical outcome for infliximab treatment in acute ulcerative colitis. *Gut* 2010; 59: 49–54.
- Paul S, Del Tedesco E, Marotte H, *et al.* Therapeutic drug monitoring of infliximab and mucosal healing in inflammatory bowel disease: a prospective study. *Inflamm Bowel Dis* 2013; 19: 2568–2576.
- Vande Casteele N, Ferrante M, Van Assche G, *et al.* Trough concentrations of infliximab guide dosing for patients with inflammatory bowel disease. *Gastroenterology* 2015; 148: 1320–1329.

14. Cornillie F, Hanauer SB, Diamond RH, *et al.* Postinduction serum infliximab trough level and decrease of C-reactive protein level are associated with durable sustained response to infliximab: a retrospective analysis of the ACCENT I trial. *Gut* 2014; 63: 1721–1727.
15. Adedokun OJ, Sandborn WJ, Feagan BG, *et al.* Association between serum concentration of infliximab and efficacy in adult patients with ulcerative colitis. *Gastroenterology* 2014; 147: 1296–307.
16. Vande Castele N, Khanna R, Levesque BG, *et al.* The relationship between infliximab concentrations, antibodies to infliximab and disease activity in Crohn's disease. *Gut* 2015; 64: 1539–1545.
17. Steenholdt C, Brynskov J, Thomsen OØ, *et al.* Individualised therapy is more cost-effective than dose intensification in patients with Crohn's disease who lose response to anti-TNF treatment: a randomised, controlled trial. *Gut* 2014; 63: 919–927.
18. Velayos FS, Kahn JG, Sandborn WJ, *et al.* A test-based strategy is more cost effective than empiric dose escalation for patients with Crohn's disease who lose responsiveness to infliximab. *Clin Gastroenterol Hepatol* 2013; 11: 654–666.
19. Steenholdt C, Ainsworth MA, Tovey M, *et al.* Comparison of techniques for monitoring infliximab and antibodies against infliximab in Crohn's disease. *Ther Drug Monit* 2013; 35: 530–538.
20. Steenholdt C, Bendtzen K, Brynskov J, *et al.* Clinical implications of measuring drug and anti-drug antibodies by different assays when optimizing infliximab treatment failure in Crohn's disease: post hoc analysis of a randomized controlled trial. *Am J Gastroenterol* 2014; 109: 1055–1064.
21. Vande Castele N, Buurman DJ, Sturkenboom MGG, *et al.* Detection of infliximab levels and anti-infliximab antibodies: a comparison of three different assays. *Aliment Pharmacol Ther* 2012; 36: 765–771.
22. Bortlik M, Duricova D, Malickova K, *et al.* Infliximab trough levels may predict sustained response to infliximab in patients with Crohn's disease. *J Crohns Colitis* 2013; 7: 736–743.
23. Pariente B, Pineton de Chambrun G, Krzysiek R, *et al.* Trough levels and antibodies to infliximab may not predict response to intensification of infliximab therapy in patients with inflammatory bowel disease. *Inflamm Bowel Dis* 2011; 18: 1199–1206.
24. Llinares-Tello F, de Salazar JRG, Gallego JMS, *et al.* Analytical and clinical evaluation of a new immunoassay for therapeutic drug monitoring of infliximab and adalimumab. *Clin Chem Lab Med* 2012; 50: 1845–1847.
25. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.
26. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
27. Vande Castele N, Herfarth H, Katz J, *et al.* American Gastroenterological Association Institute technical review on the role of therapeutic drug monitoring in the management of inflammatory bowel diseases. *Gastroenterology* 2017; 153: 835–857.
28. Afonso J, Lopes S, Gonçalves R, *et al.* Proactive therapeutic drug monitoring of infliximab: a comparative study of a new point-of-care quantitative test with two established ELISA assays. *Aliment Pharmacol Ther* 2016; 44: 684–692.
29. Moore C, Corbett G and Moss AC. Systematic review and meta-analysis: serum infliximab levels during maintenance therapy and outcomes in inflammatory bowel disease. *J Crohns Colitis* 2016; 10: 619–625.
30. Ruiz-Argüello B, del Agua AR, Torres N, *et al.* Comparison study of two commercially available methods for the determination of infliximab, adalimumab, etanercept and anti-drug antibody levels. *Clin Chem Lab Med* 2013; 51: 287–289.
31. Schmitz EM, van de Kerkhof D, Hamann D, *et al.* Therapeutic drug monitoring of infliximab: performance evaluation of three commercial ELISA kits. *Clin Chem Lab Med* 2016; 54: 1211–1219.
32. van Bezooijen JS, Koch B, van Doorn M, *et al.* Comparison of three assays to quantify infliximab, adalimumab, and etanercept serum concentrations. *Ther Drug Monit* 2016; 38: 432–438.
33. Afonso J, Lopes S, Gonçalves R, *et al.* Detection of anti-infliximab antibodies is impacted by antibody titer, infliximab level and IgG4 antibodies: a systematic comparison of three different assays. *Therap Adv Gastroenterol* 2016; 9: 781–794.