Systematic Review of the Performance of HIV Viral Load Technologies on Plasma Samples

Kimberly A. Sollis¹, Pieter W. Smit¹, Susan Fiscus², Nathan Ford³, Marco Vitoria³, Shaffiq Essajee⁴, David Barnett⁵, Ben Cheng⁶, Suzanne M. Crowe⁷, Thomas Denny⁸, Alan Landay⁹, Wendy Stevens¹⁰, Vincent Habiyambere¹¹, Jos Perrins¹¹, Rosanna W. Peeling¹∗

¹ Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, United Kingdom, ² Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, North Carolina, United States of America, ³ Department of HIV/AIDS, World Health Organization, Geneva, Switzerland, ⁴ HIV, Medicine and Science, Clinton Health Access Initiative, New York, New York, United States of America, ⁵ Department of Haematology, United Kingdom National External Quality Assessment Service (UK NEQAS) for Leucocyte Immunophenotyping, Sheffield, United Kingdom, ⁶ Department of Technology and Innovation, Pangaea Global AIDS Foundation, San Francisco, California, United States of America, ⁷ Centre for Biomedical Research, Burnet Institute, Melbourne, Australia, ⁸ Department of Medicine, Duke Human Vaccine Institute and Center for HIV/AIDS Vaccine Immunology, Durham, North Carolina, United States of America, ⁹ Department of Immunology-Microbiology, Rush University Medical Center, Chicago, Illinois, United States of America, ¹⁰ Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa, ¹¹ Department of HIV/AIDS, World Health Organization, Geneva, Switzerland

Abstract

Background: Viral load (VL) monitoring is the standard of care in developing country settings for detecting HIV treatment failure. Since 2010 the World Health Organization has recommended a phase-in approach to VL monitoring in resource-limited settings. We conducted a systematic review of the accuracy and precision of HIV VL technologies for treatment monitoring.

Methods and Findings: A search of Medline and Embase was conducted for studies evaluating the accuracy or reproducibility of commercially available HIV VL assays. 37 studies were included for review including evaluations of the Amplicor Monitor HIV-1 v1.5 (n = 25), Cobas TaqMan v2.0 (n = 11), Abbott RealTime HIV-1 (n = 23), Versant HIV-1 RNA bDNA 3.0 (n = 15), Versant HIV-1 RNA kPCR 1.0 (n = 2), ExaVir Load v3 (n = 2), and NucliSens EasyQ v2.0 (n = 1). All currently available HIV VL assays are of sufficient sensitivity to detect plasma virus levels at a lower detection limit of 1,000 copies/mL. Bias data comparing the Abbott RealTime HIV-1, TaqMan v2.0 to the Amplicor Monitor v1.5 showed a tendency of the Abbott RealTime HIV-1 to under-estimate results while the TaqMan v2.0 overestimated VL counts. Compared to the Amplicor Monitor v1.5, 2–26% and 9–70% of results from the Versant bDNA 3.0 and Abbott RealTime HIV-1 differed by greater than 0.5log₁₀. The average intra and inter-assay variation of the Abbott RealTime HIV-1 were 2.95% (range 2.0–5.1%) and 5.44% (range 1.17–30.00%) across the range of VL counts (2log₁₀–7log₁₀).

Conclusions: This review found that all currently available HIV VL assays are of sufficient sensitivity to detect plasma VL of 1,000 copies/mL as a threshold to initiate investigations of treatment adherence or possible treatment failure. Sources of variability between VL assays include differences in technology platform, plasma input volume, and ability to detect HIV-1 subtypes. Monitoring of individual patients should be performed on the same technology platform to ensure appropriate interpretation of changes in VL. Prospero registration #: CD42013003603.

Introduction

As of mid 2013 it is estimated that over nine million HIV infected individuals are on antiretroviral therapy (ART) worldwide and a substantial proportion have been on treatment for ten years or more [1]. As the global ART cohort continues to expand and mature, the need for ongoing monitoring is becoming increasingly important to ensure treatment efficacy and minimize the risk of HIV drug resistance. Clinical and immunological monitoring techniques have poor sensitivity and specificity for detecting virologic failure, leading to a substantial misclassification of treatment responses, resulting in delayed switching in some cases and inappropriate switching from first line regimens in others [2–7].

Routine HIV viral load (VL) monitoring has the potential to improve the accuracy of diagnosis of treatment failure, enable more targeted adherence interventions, and preserve the efficacy of ART [8]. Monitoring HIV VL is often not performed in resource-limited settings because the assays are costly, and require sophisticated, expensive laboratory equipment and trained technicians [9,10]. Despite these limitations, the importance of HIV
VL testing is increasingly recognized; in 2010 the World Health Organization (WHO) recommended that countries begin to phase in VL for monitoring patients on ART [1], a recommendation reinforced in the 2013 treatment guidelines [11]. Detailed descriptions of available VL technologies can be found in a UNITAID HIV/AIDS diagnostic landscaping [12]. In order to support decisions regarding which VL tools to phase in, we conducted a systematic review of the performance and operational characteristics of commercially available HIV VL assays.

**Methods**

We first verified that no systematic reviews had already been conducted on this topic by searching the Cochrane Library and Centre for Reviews and Dissemination, University of York and National Institute for Health Research. A research protocol was then developed following standard guidance [13] and this was reviewed by all members of the HIV Monitoring Technologies Working Group before the search was performed. The systematic review protocol was registered with PROSPERO (http://www.crd.york.ac.uk/PROSPERO), registration number CD42013003603.

**Search**

Medline and Embase were searched using the search terms ‘HIV-1’ or ‘HIV-2’ or ‘HIV’ or ‘human immunodeficiency virus’ or ‘HIV type 1’ or ‘HIV type 2’ or ‘human immunodeficiency virus type 1’ or ‘human immunodeficiency virus type 2’ and ‘viral load’ or ‘viral RNA’ and ‘(compar*) or (eval*)’ and ‘(measure* or quant* or techn* or test*)’ and ‘(accuracy or performance or precision or ‘sensitivity or specificity)’. Results of the search were exported to EndNote X3, duplicates removed and the remainder assessed for relevance and fulfillment of the selection criteria.

**Study Selection**

The search was conducted in February 2010 and updated in April 2012 to include scientific research articles published in peer-reviewed journals, in English, between January 1990 and the search date. Publications evaluating or comparing the performance of commercial assays for the quantification of HIV-1 or HIV-2 virus load in plasma were included in the search. There were no limitations on the method of nucleic acid extraction, amplification, or detection but the assays under investigation had to be commercially available at the time of the review. The study population was limited to adults but no restriction was placed on the geographical origin of the samples or the HIV subtype (HIV-1 or HIV-2). Publications using samples from standardized panels were also considered for inclusion providing they met the study criteria. No authors were contacted for further information and all data presented in this review were available in the included publications.

**Data collection processes**

Two independent reviewers extracted data on assay accuracy and reproducibility from publications meeting the inclusion criteria as defined in the protocol. Where there was any discrepancy, the reviewers met to discuss the difference and came to a consensus on inclusion or exclusion from the study. The quality of publications included in the HIV VL review was scored using adapted STARD guidelines [14,15]. This included questions on the title and abstract; introduction; methods including participant/sample characteristics, test methods and statistical methods; results including data on participants and test results; and discussion (Annex S3). The two reviewers selected 17 critical quality criteria of the original 23 which were more appropriate for evaluations of quantitative assays.

**Quantitative data synthesis**

Accuracy and reproducibility data were summarized graphically in Excel. Accuracy measures included bias and limits of agreement [16], sensitivity and specificity, and the percentage of results differing by 0.5Log_{10}, which is generally considered the clinically relevant difference between two VL measurements [17,18]. Reproducibility measures included within- and between-assay variability, reported as % coefficient of variation (%CV).

**Results**

**Study selection**

The search produced 1,715 titles, of which 580 were removed as duplicates. Of the remaining 1,135 titles and abstracts, 261 publications were reviewed as full text, and 37 met the criteria for inclusion and were taken forward for inclusion in the review (Figure 1) [19–55].

**Study characteristics**

The studies included data on the following assays: Roche Amplicor Monitor v1.5 (Amplicor 1.5) (n = 23), Roche Cobas TaqMan v2.0 (TaqMan v2.0) (n = 11), Abbott RealTime HIV-1 (Abbott RealTime) (n = 23), Versant HIV-1 RNA bDNA 3.0 (bDNA 3.0) (n = 15), Versant HIV-1 RNA kPCR 1.0 (kPCR 1.0) (n = 2), Cavidi ExaVir Load v3 (ExaVir v3) (n = 2) and the NucliSens EasyQ 2.0 (EasyQ 2.0) (n = 1). Only one publication evaluated assays using HIV-2 samples (29) and the remainder used HIV-1 Group M, N, and O. There was no single standard reference test used as a comparator in all studies. Of the 37 included studies, 12 were published in 2009 (median), with a minimum and maximum of 2000 and 2012 respectively.

**Quantitative Data Synthesis: Accuracy of HIV VL Assays**

**Analytical Sensitivity and Specificity of HIV VL Assays with Plasma.** One study provided analytical sensitivity and specificity data for ExaVir Load v3 compared to Amplicor Monitor v1.5 [28] (Table 1). Sensitivity of the ExaVir v3 was 96–100% at HIV VL concentrations above 2000 copies/mL, but decreases to 59% when VL concentration decreased to between 50–400 copies/mL. The specificity of the ExaVir v3 was evaluated using HIV-1 negative samples and reported as 100% [26].

Five studies evaluated the specificity of the Amplicor v1.5, Abbott RealTime, bDNA 3.0, and ExaVir v3 assays using HIV-1 negative samples and reported as 100% [15,18,28,45,49]. When tested with a panel containing four HIV-2, four HCV, and four polyomavirus BK plasma samples, the specificity of the kPCR was 92% [39]. One study evaluated the specificity of the TaqMan v2.0 using HIV-1 negative samples containing potentially cross-reactive reactive viruses (including adenovirus Type 5, cytomegalovirus, Epstein-Barr virus, hepatitis B, C, and A viruses, herpes simplex virus Type I and Type II and others (n = 660) and found no false positive results or cross-reactivity [44].

**Clinically important differences in result.** Seventeen studies evaluated the percentage of results differing by a clinically important value, defined as 0.5Log_{10}, between the index test and reference test [22,27–29,31,38–41,43,48–51,55]. Data were available for Amplicor v1.5, TaqMan v2, Abbott RealTime, bDNA 3.0, kPCR, and ExaVir v3.
Between 8.5% [49] and 70.0% [38] of results provided by the Abbott RealTime assay differed by greater than 0.5log_{10} compared to the Roche Monitor v1.5 (Figure 2). The greatest differences in results occurred using the 1 mL Abbott RealTime sample input, where 70% of results differed by more than 0.5log_{10} compared to the Amplicor 1.5 [38]. Results from the bDNA 3.0 showed much lower levels of discordance compared to the Amplicor 1.5, with between 2–26% of results having clinically
important differences [22,27]. Only one study reported differences between results from the ExaVir v3.0 and the Amplicor v1.5; in this study, 27% of results from the ExaVir differed by more than 0.5log<sub>10</sub> [28].

When comparing the assays to the Abbott RealTime assay, between 10 and 30% of TaqMan results [41,51], and 13% of bDNA v3 results [48] differed by more than 0.5log<sub>10</sub>. Bias. Bias was measured between an index test and the Amplicor v1.5 as a reference standard in 14 publications [19,24,25,29,37,38,40,41,43,45,46,48,52,55]. These included comparisons with the Abbott RealTime (n = 11) [19,24,25,29,38,40,45,46,48,55], TaqMan v2.0 (n = 4) [24,25,41], and bDNA 3.0 (n = 4) [37,43,52]. There were no bias data available comparing the ExaVir v3, EasyQ v2.0, or the kPCR to the Monitor v1.5. Figure 3 summarizes the range of bias measurements reported and the variability in the limits of agreement reported for each assay when evaluated alongside the Amplicor v1.5.

The EasyQ reports results as IU/mL (International Units/mL). A conversion factor supplied by the manufacturer was applied to enable comparison with other studies; however this process did not produce consistent results when applied to the limits of agreement.

Data on bias were also available comparing index tests to the Abbott RealTime (Figure 4) [19,29,30,32,35,41,42,45,48,53], TaqMan v2.0 (Figure 5) [24,30,35,41,42,50,51,54] and to a WHO International Standard (Table 2) [51]. Compared to the Abbott RealTime, the Taqman v2.0 overestimated VL counts by 0.04–0.33 log<sub>10</sub> c/mL [30,35,41,42], while the ExaVir v3 and kPCR underestimated VL but only by 0.28 and 0.16 log<sub>10</sub>/mL, respectively [32,39]. The limits of agreement associated with the ExaVir v3 spanned from −1.27 to 0.72 log<sub>10</sub>/mL [32], and those associated with the kPCR data point from −0.474 to 0.154 log<sub>10</sub>/mL [39]. When index tests were compared to the TaqMan v2.0, the kPCR underestimated VL [50]. The Abbott RealTime both over and under-estimated HIV VL values but never by more than 0.5log<sub>10</sub> [30,35,41,42,51], with the exception of one study that reported limits of agreement spanning from −1 to 0.6 log<sub>10</sub>/mL [42]. EasyQ 2.0 overestimated VL by 0.88log<sub>10</sub> c/mL [54]. Finally, when compared to the WHO International Standard, neither the Abbott RealTime nor the TaqMan v2.0 differed by more than 0.5log<sub>10</sub> with the Abbott RealTime displaying slight overestimation while the TaqMan v2.0 showed slight underestimation of HIV VL [51].

Quantitative Data Synthesis: Precision of HIV VL Assays

Data on intra-assay (within-run) and inter-assay (between-run) variability were reported for six assays: Amplicor v1.5 [23,47], Abbott RealTime (Figure 6) [19,20,23,38,40,45], and the kPCR (Figure 7) [39,50]. The Abbott RealTime showed excellent intra- and inter-assay reproducibility (<10% variability) at low copy

---

Table 1. Analytical Sensitivity of the ExaVir Load v3 compared to Amplicor Monitor v1.5.

<table>
<thead>
<tr>
<th>Cut-off (copies/mL)</th>
<th>Publication</th>
<th>N</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;200 (28)</td>
<td></td>
<td>199</td>
<td>92%</td>
</tr>
<tr>
<td>&gt;400 (28)</td>
<td></td>
<td>178</td>
<td>94%</td>
</tr>
<tr>
<td>&gt;1000 (28)</td>
<td></td>
<td>145</td>
<td>98%</td>
</tr>
<tr>
<td>&gt;5000 (28)</td>
<td></td>
<td>23</td>
<td>100%</td>
</tr>
<tr>
<td>50–400 (28)</td>
<td></td>
<td>64</td>
<td>59%</td>
</tr>
<tr>
<td>401–2000 (28)</td>
<td></td>
<td>57</td>
<td>86%</td>
</tr>
<tr>
<td>2001–10 000 (28)</td>
<td></td>
<td>50</td>
<td>96%</td>
</tr>
<tr>
<td>10 001–50 000 (28)</td>
<td></td>
<td>48</td>
<td>100%</td>
</tr>
</tbody>
</table>

---

Figure 2. % of results differing by >0.5log10 between index test and Amplicor v1.5 (data extracted from references [22,27–29,38,40,45,48,55]). *No volume specified.

doi:10.1371/journal.pone.0085869.g002
Figure 3. Bias between Index Test and Ampicor v1.5 as a comparator (data extracted from references [19,24,25,29,37,38,40,41,43,45,46,48,52,55]). *No volume specified.
doi:10.1371/journal.pone.0085869.g003

Figure 4. Bias between Index Test and Abbott RealTime as a comparator (data extracted from references [19,30,32,35,41,42,48]).
doi:10.1371/journal.pone.0085869.g004
numbers. One study, however, found the inter-assay variability to be 30% at 500 000 c/mL [19]. The kPCR showed overall intra- and inter-assay variability exceeding 15% [39,50].

Quality Assessment of Studies Included in the HIV VL Review

All thirty-seven articles included in the review were assessed for quality by two independent reviewers (Annex S1, S2). No article met all 17 quality assessment criteria. The quality scores ranged from 24–94%, and the median was 65%. While 95% of articles described the study aims, only 8% reported on staff training. Twenty-three (62%) and twenty-six (70%) of included publications clearly described sample acquisition and sample storage conditions, respectively. Twenty-one studies (57%) detailed the statistics performed but only 16 (43%) presented descriptive statistics and bias calculations. All studies discussed the clinical relevance of their findings.

Discussion

In 2013, the World Health Organization recommended that, with the exception of dried blood spot samples, the threshold for detection of virological failure should be lowered to 1000 c/mL [11]. This recommendation was made in support of a shift towards a more strategic use of antiretrovirals both for the treatment of HIV infection and also the prevention of onward transmission through earlier initiation of ART among priority groups such as pregnant women and serodiscordant couples [11].

This review found that all the assays currently in use can reliably detect HIV VL of 1000 c/mL, which is within the linear ranges of VL assays claimed by manufacturers (Table 3). If a threshold of ≥1000 c/mL is used to consider switching to a second line regimen, then all assays were found to have acceptable performance to be of use in clinical decision making. The challenge of routinely and reliably detecting 1000 c/mL may be of greater concern for the next generation of point-of-care tests.

The most difficult aspect of conducting this review was the different reference standards used for evaluating test performance in different studies. For comparability it would be useful if a single standard measurement was used for HIV VL. The NucliSens EasyQ is the only assay to report results using IU/mL. As technologies evolve, a consensus international standard for HIV VL copies that is widely accessible would provide a valid and easier reference standard for determining the analytical performance of a new assay.

Sources of variability between VL assays reported include not only differences in technology platform, but also plasma input volume, and ability to detect HIV-1 subtypes. VL monitoring should therefore be performed on the same technology platform for monitoring individual patients to ensure appropriate interpretation of changes in VL, unless clinically relevant differences are not identified between different assays.

Table 2. Bias data between index test and WHO International Standard [50].

<table>
<thead>
<tr>
<th>Index Test</th>
<th>Reference Standard (mean log copies/ml)</th>
<th>Bias</th>
<th>Limits Of Agreement (min)</th>
<th>Limits Of Agreement (max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott RealTime HIV-1</td>
<td>WHO International Standard Low (2.29)</td>
<td>−0.3</td>
<td>−0.44</td>
<td>−0.16</td>
</tr>
<tr>
<td>Abbott RealTime HIV-1</td>
<td>WHO International Standard High (4.38)</td>
<td>−0.26</td>
<td>−0.31</td>
<td>−0.21</td>
</tr>
<tr>
<td>Roche TaqMan v2.0</td>
<td>WHO International Standard Low (2.30)</td>
<td>0.31</td>
<td>0.22</td>
<td>0.39</td>
</tr>
<tr>
<td>Roche TaqMan v2.0</td>
<td>WHO International Standard High (4.40)</td>
<td>0.17</td>
<td>0.1</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Figure 5. Bias between Index Test and TaqMan v2.0 as a comparator (data extracted from references [30,35,41,42,50,51,54]). *No volume specified.

doi:10.1371/journal.pone.0085869.g005
Interpretation of the data available was also limited by the variable quality of the publications. This review highlights the need for more rigor in the design and reporting of evaluations of HIV VL quantification technologies, particularly as new versions of HIV VL assays and point-of-care (POC) formats become available. One shortcoming highlighted by the review is the incorrect application of statistical techniques. Correlation and linear regression were the most common measurements reported but bias and limits of agreement would be much more informative. Unlike linear regression, Bland-Altman plots describe the mean difference between two sets of data points and give this value a direction indicating whether the index test is likely to under- or over-quantify results [16]. As with CD4 quantification technologies, the extent of misclassification above and below a clinically important threshold will need to be investigated [36]. It is important for future studies to report the frequency, intensity and direction of the misclassifications [1], because misclassification can have clinical and public health implications (patients are left on failing regimens and may develop drug resistance) and economic implications (second line regimens are often expensive and options beyond second-line are limited). Precision or reproducibility should also be detailed with a clear description of how the measures were obtained, including information on number of samples, number of replicates per sample, and a descriptive summary of the characteristics of the samples used including mean HIV VL (± SD) and range. The results of the review show that standardized practices and guidelines for improved methods undertaking and reporting evaluations of HIV VL assay evaluations are needed, particularly with respect to defining the study population, reporting algorithms for inclusion and exclusion of samples throughout the study, reporting training of technicians, and the use of appropriate statistical methods [11].

The main limitation of this review methodology is that the inclusion criteria were limited to studies published in English, which may have overlooked useful data available in other languages.

Since the results of our review indicate that all currently commercially available HIV VL assays can provide a reliably accurate measure of plasma VL ≥1000 c/mL, switching from the current WHO recommended threshold of 5,000 c/mL for investigations for treatment compliance or possible treatment failure to 1,000 c/mL would allow earlier detection of treatment failure, enable more targeted adherence interventions, and preserve the efficacy of ART. Choice of technology platform should take into account the ability to detect HIV-1 subtypes in the target population. Serial samples for VL monitoring need to be performed on the same technology platform for proper interpretation of any meaningful changes in VL.

![Figure 6. Intra- and inter-assay variation for the Abbott RealTime HIV-1(plasma) according to log copy number/mL of sample (data extracted from references [19,20,23,38,40,45]). doi:10.1371/journal.pone.0085869.g006](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAgAAAAAgCAYAAABGemrpAAAABGdBTUEAALGPCgorithmil.png)

![Figure 7. Intra- and inter-assay variation for the Versant kPCR (plasma) according to log copy number/mL of sample (data extracted from references [39,50]). doi:10.1371/journal.pone.0085869.g007](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAgAAAAAgCAYAAABGemrpAAAABGdBTUEAALGPCgorithmil.png)

### Table 3. Manufacturer Data: Plasma Input Volume and Linear Range of Detection.

<table>
<thead>
<tr>
<th>Manufacturer: Assay</th>
<th>Input vol</th>
<th>Linear Range (c/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott RealTime HIV-1</td>
<td>0.2 mL</td>
<td>0.2 mL sample input: 150–10,000,000</td>
</tr>
<tr>
<td></td>
<td>0.5 mL</td>
<td>0.5 mL sample input: 75–10,000,000</td>
</tr>
<tr>
<td></td>
<td>0.6 mL</td>
<td>0.6 mL sample input: 40–10,000,000</td>
</tr>
<tr>
<td></td>
<td>1.0 mL</td>
<td>1.0 mL sample input: 40–10,000,000</td>
</tr>
<tr>
<td>bioMérieux NucliSens EasyQ HIV-1 v2.0</td>
<td>0.1 mL</td>
<td>10–10,000,000</td>
</tr>
<tr>
<td>CaviDi ExaVir Load</td>
<td>1 mL</td>
<td>200–600,000</td>
</tr>
<tr>
<td>Roche Diagnostics Cobas Amplicor HIV-1 MONITOR TEST, v1.5</td>
<td>0.5 mL</td>
<td>Standard protocol: 400–750,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ultra-sensitive protocol: 50–100,000</td>
</tr>
<tr>
<td>Roche COBAS AmpliPre/COBAS TaqMan HIV-1 Test, v2.0</td>
<td>1 mL</td>
<td>20–100,000,000</td>
</tr>
<tr>
<td>Siemens VERSANT HIV-1 RNA 3.0 assay (bDNA)</td>
<td>1.0 mL</td>
<td>50–500,000</td>
</tr>
<tr>
<td>Siemens VERSANT HIV-1 RNA 1.0 assay (kPCR)</td>
<td>0.65 mL</td>
<td>37–11,000,000</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0085869.t003
Annex S1  Quality assessment results by 17 criteria.


Annex S3  PRISMA Guidelines.