Performance of the Xpert® HIV-1 Viral Load assay: A systematic review and meta-analysis

Madlen Nash1,2, Sophie Huddart1,2, Sayema Badar3, Shrikala Baliga4,5, Kavitha Saravu5,6, Madhukar Pai1,2, #

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
2 McGill International TB Centre, McGill University, Montreal, Canada
3 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, Tulane, New Orleans
4 Department of Microbiology, Kasturba Medical College, Mangalore, Manipal University, Mangalore, India
5 Manipal McGill Center for Infectious Diseases, Manipal University, Manipal, India
6 Department of Medicine, Kasturba Medical College, Manipal, Manipal University, Mangalore, India

Running Head: Performance of the Xpert® HIV-1 Viral Load assay

Correspondence

Prof Madhukar Pai, MD, PhD, FCAHS
Canada Research Chair in Epidemiology & Global Health
McGill University
Dept of Epidemiology & Biostatistics
1020 Pine Ave West
Montreal, QC H3A 1A2, Canada
Tel: 514-398-5422
Email: madhukar.pai@mcgill.ca
Abstract

Viral load (VL) is the preferred treatment monitoring approach for HIV-positive patients. However, more rapid, near-patient, and low-complexity assays are needed to scale-up VL testing. The Xpert HIV-1 VL assay (Cepheid, Sunnyvale) is a new, automated molecular test, and can leverage the GeneXpert systems that are being used widely for tuberculosis diagnosis. We systematically reviewed the evidence on the performance of this new tool in comparison to established reference standards. A total of twelve articles (thirteen studies) in which HIV patient VLs were compared between Xpert HIV VL assay and a reference standard VL assay were identified. Study quality was generally high but substantial variability was observed in the number and type of agreement measures reported. Correlation coefficients between Xpert and reference assays were high with a pooled Pearson correlation (n=8) of 0.94 [0.89, 0.97] and Spearman correlation (n=3) of 0.96 [0.86, 0.99]. Bland-Altman metrics (n=11) were all within 0.35 log copies/mL of perfect agreement. Overall, Xpert HIV-1 VL performed well in comparison with current reference tests. The minimal training and infrastructure requirements for the Xpert HIV-1 VL assay make it attractive for use in resource constrained settings, where point-of-care VL testing is most needed.
Introduction

Despite the recommendations put forward by the World Health Organization (WHO), viral load (VL) monitoring of antiretroviral therapy (ART) is not routinely performed in many low-resource countries and treatment failure is diagnosed on the basis of clinical or immunological criteria. Currently used VL assays demand sophisticated facilities, expensive equipment and skilled technicians, making them unaffordable and largely impractical for scale-up in resource-limited settings (1). To expand the use of targeted and routine VL monitoring, inexpensive, low-complexity assays are needed, preferably for point-of-care use.

The Xpert HIV-1 VL assay (Cepheid, Sunnyvale, CA, USA), performed on the GeneXpert® Instrument System, is an in vitro diagnostic test designed for the rapid quantification of HIV-1 in human plasma from individuals with an active HIV infection. It uses Real-Time Quantitative Reverse Transcription PCR technology, and targets HIV-1 Group M subtypes A, B, C, D, AE, F, G, H, AB, AG, J, K and Group N and Group O. It has a limit of quantitation of 40 cp/mL and can detect HIV-1 RNA over a linear range of 40 to 10,000,000 copies/ml. The GeneXpert platform allows on-demand molecular testing in one fully integrated closed-cartridge and provides results in 90 minutes (2). The Xpert HIV-1 VL assay runs on the same GeneXpert platform as the WHO-endorsed Xpert MTB/RIF cartridge, used for diagnosis of tuberculosis.

The Xpert MTB/RIF assay is a major advance in tuberculosis diagnostics, and over 23 million cartridges have been used around the world (3, 4). Given the high co-prevalence of TB and HIV in many settings (5), and need for greater integration of TB-HIV services, leveraging the existing GeneXpert network for HIV may substantially increase access to VL testing. Similar to the Xpert MTB/RIF assay, the Xpert HIV-1 VL assay demands minimal training and modest infrastructure requirements while providing rapid
If proven to be as accurate as current, established reference standard VL tests, this monitoring tool has the potential for rapid scale-up in countries already using the GeneXpert platform for TB.

Scaling up routine HIV-1 VL testing is essential to meet the UNAIDS 90-90-90 targets.

In July of 2017, the Xpert HIV-1 VL was accepted for the WHO list of prequalified in vitro diagnostics. Country level validation studies on the accuracy of this assay have now been conducted in a variety of settings. We conducted a systematic review to synthesize evidence from these validation studies, and estimate the overall agreement between the Xpert HIV-1 VL assay and current reference standard assays.

Methods

We conducted a systematic review of the literature according to PRISMA guidelines. A written protocol for the systematic review was prepared a priori. A modification to the pooling criteria was made after the authors were made aware of I²’s tendency to approach 100% as sample size of studies increases. Thus, for the pooling of correlations, where internal study variability was very small and resulting I² was very high, the decision to pool was based on clinical heterogeneity.

Search strategy and selection criteria

Using OVID, we systematically searched Medline (1946-2017), Embase (1947-2017) and Global Health (1973-2017) for studies evaluating the performance of the Xpert HIV-1 VL assay in HIV-positive patients (see appendix for complete search strategy). No language or geographic restriction was applied. The study search was conducted on September 25th, 2017.

All studies that enrolled participants with a known HIV-positive status were eligible for inclusion.
We restricted inclusion to studies of patients of known HIV status because VL testing is not recommended as a HIV diagnostic but as a treatment monitoring tool. The included studies were required to compare Cepheid’s Xpert HIV-1 VL assay to another established PCR based VL assay and report at least one comparison measure between Xpert VLS and gold standard VLS (i.e., Pearson or Spearman correlation, Bland-Altman agreement etc.). Conference abstracts and posters were excluded as their quality could not be comprehensively assessed (Figure 1).

Two reviewers (SH and MN) screened the titles and abstracts of citations retrieved from all sources. Duplicates were removed and studies which met the inclusion criteria were flagged for further review. A full text screen of relevant studies was then performed by the same independent reviewers.

Data extraction
Two reviewers (SH and MN) independently extracted data from all eligible studies on patient demographics, correlation coefficients and results of Bland-Altman analyses comparing Xpert HIV-1 VL assays and the reference test and study quality (see appendix).

Extracted data was adjudicated by the two reviewers with discrepancies resolved by a third reviewer (SB). Study authors were contacted regarding information not reported in the included manuscripts but no responses were received.

Risk of bias and assessment of study quality
A modified version of the QUADAS-2 criteria for diagnostic tests was used to assess study quality (10). Because of the numerical output of both the index and reference tests, biased interpretation of either assay was unlikely and we thus did not evaluate blinding of the index test or reference test as sources of
We assessed patient selection and patient flow according to QUADAS-2 guidelines. Overall risk of bias for patient selection and flow was summarized as low, high or unclear.

**Statistical analysis**

Correlation coefficients, both Spearman and Pearson, were extracted from studies reporting these measures. Because of the theoretical heterogeneity between parametric and nonparametric measures, meta-analysis was only considered within strata of correlation type. The decision to pool was based on the author’s assessment of clinical heterogeneity.

The results of Bland-Altman analyses, a summary of the mean differences between VL measures, were also extracted. Twelve Bland-Altman analyses were performed among the included studies but one (11) did not report the corresponding standard deviation, confidence interval or limit of agreement. Without a measure of variance, it is difficult to appropriately interpret this Bland-Altman analysis thus the study’s Bland-Altman analysis was not included in the quantitative summary. The decision to pool for Bland-Altman was based on $I^2$ as internal study variability was moderate and $I^2$ appropriately reflected heterogeneity.

All extracted studies performed their correlation and Bland-Altman analyses on log-transformed viral loads. For studies on patient samples, samples were independent with one sample per patient included in the agreement analyses. For one included study using lab quality assessment samples the sampling procedure was not described (12) (lab samples).

Meta-analysis was performed using a DerSimonian and Laird random effects model (13).
The number of studies was insufficient to conduct meta-regression. Traditional methods to assess publication bias such as Egger’s test or funnel plots are severely underpowered especially with small numbers of studies thus publication bias was not formally assessed but it is assumed to exist to some degree in all systematic reviews.

All statistical analyses were performed in R.

Results

Figure 1 shows the study selection flow chart. Twelve articles covering 13 studies were identified during the systematic search of the literature; these studies enrolled more than 3,300 individual patients, 2,011 of whom had VLs quantifiable on Xpert and the reference test (Table 1) (11, 12, 14-23). For most studies, patient blood samples were collected as part of routine clinical practice either within a larger HIV/AIDS study or through healthcare institutions. One paper included data on the quality assessment samples (12). Three studies enrolled HIV patients in India (20, 22, 23) and three studies enrolled HIV patients in South Africa (12, 14) (Gous, patient samples). Overall, the quality of reporting of patient demographics was poor. Only three of thirteen studies reported a complete set of demographic covariates (average patient age, gender distribution and the proportion of participants receiving ART) (12, 14, 20). Of those studies that reported ART coverage, treatment rates ranged from 0 to 100%. The reference standards used included PCR-based VL assays (Abbott RealTime HIV-1 m2000rt, Abbott Molecular; COBAS® AmpliPrep / COBAS® TaqMan®, Roche Diagnostics; NuclISENS EasyQ® HIV-1 v2.0, bioMérieux; and VERSANT HIV-1 RNA 1.5 Assay, Siemens HealthCare Diagnostics).

Overall, when pertinent information was reported, study quality was high (Table 2). Validation studies require cross-sectional design which minimizes risk of potential biases. Risk of selection bias, non-
generalizability or risk of flow bias was generally low. Some studies did not clearly state aspects related to study design (i.e., sequential or random enrollment, reasons for exclusion from analysis) which limits the ability to draw conclusions about study quality.

All reported correlations between the Xpert HIV-1 VL assay and gold standard tests were very high for both Pearson and Spearman correlation coefficients (Figure 2). As all correlations within strata were very close, ranging from 0.81 to 0.98 studies were pooled within strata of correlation type. The pooled correlations were 0.94 [0.89, 0.97] and 0.96 [0.86, 0.99] for Pearson and Spearman correlations, respectively. These values indicate a very high degree of agreement between Xpert VL and reference standard VL values.

Bland-Altman results are normally distributed thus $I^2$ is an appropriate metric for heterogeneity. The $I^2$ for the Bland-Altman results was 96% which suggests enough heterogeneity to preclude pooling. However, all studies reported a mean difference within 0.34 units of zero (the ideal value) (Figure 3). We stratified the correlation and Bland-Altman values by ART status and found no major difference in studies with patients on ART versus studies with ART-naïve or mixed ART status patients (Figures S1 and S2).

There were a sufficient number of Bland-Altman studies to stratify by type of reference test though heterogeneity remained too high to provide pooled estimates (Figure S3). No major difference in Bland-Altman results by reference test were identified.

We observed substantial variability in the number and type of agreement measures reported in the included studies (Table 3). Correlation, either Pearson or Spearman, and Bland-Altman measures were the most commonly reported but several other measures were reported. Additionally, one study (11)
reported a mean log difference as part of its Bland-Altman analysis but did not provide a standard deviations, confidence intervals, or limits of agreement. This prevented some studies from being included in the Bland-Altman meta-analysis.

Discussion

Overall, our systematic review showed Xpert HIV-1 VL performs well in comparison with current established reference standard VL assays, both when measured by correlation and by Bland-Altman analysis. These findings might help inform policy guidance on this new assay, along with data on costs, feasibility, clinical impact, and cost-effectiveness.

Our systematic review has several potential limitations. First, as the search was conducted shortly after Xpert HIV-1 VL received WHO in vitro diagnostic prequalification, few validation studies had been conducted and published at the time of the search. Consequently, only twelve articles meeting our inclusion criteria were identified and included. We hope to update our meta-analysis in future, to account for new studies that will emerge on the Xpert VL assay. Second, the decision to exclude conference abstracts resulted in the exclusion of multiple relevant studies. Abstracts were excluded as study quality could not be effectively assessed, but as Xpert HIV-1 VL is a new tool, many studies on the assay have yet to be published as full-length articles. However, due to poor reporting, study quality was challenging to assess even in full-length articles. Where sufficient information was available, quality was high.

The final limitation pertains to generalizability. Many studies failed to report important demographic and clinical information such as gender distribution, age of study participants and patient ART status. This information is critical to understand the clinical relevance and generalizability of the reported data.
Authors of studies with missing information were contacted, however, no responses were received.

There was also substantial variability in which measures of agreement were calculated. To improve the ability to systematically review and pool the literature, a minimum set of standard measures of agreement should be agreed upon and used. Correlation and Bland-Altman could provide such a minimum set as they measure both general agreement and magnitude and direction of bias between assays. Many studies reported percent agreement analyses where Xpert HIV-1 VL and the reference test were used to classify patients above or below certain thresholds. However, the specific thresholds used varied considerably preventing a meaningful comparison. The use of a standard clinical threshold in future evaluations of Xpert HIV-1 VL would allow for direct comparisons in systematic reviews. The generalizability of our study is also limited due to the geographical distribution of the primary studies included; they were exclusively performed in countries where the predominate HIV subtype is C (e.g., India and Africa) or B (e.g., Israel, Europe and the United States) (24). As more validation studies continue to be published on the Xpert assay, the results of this meta-analysis can be updated to include data from different countries where other HIV-1 subtypes predominate.

The purpose of this study was to assess VL agreement between Xpert and other gold standard assays as a measure of accuracy. However, a final recommendation for the use of Xpert VL will need to consider not just accuracy, but country-specific implementation research that addresses the feasibility of Xpert VL and its impact on patient outcomes.

The low-complexity nature of the Xpert HIV-1 VL assay coupled with its ability to deliver same day test results make this technology particularly well suited for use in resource constrained settings, where point-of-care VL testing is most needed (6). Deployment of such a tool as well as utilization of pre-existing Gene Xpert systems which are used in TB diagnostics has the potential to increase access to VL.
testing, which will be necessary to achieve the 90-90-90 global targets for HIV/AIDS. Further research is needed to assess the impact of this VL assay on patient-important outcomes, and to establish its cost-effectiveness compared to currently used VL assays in centralized laboratories.
Acknowledgements

This study was supported by a TMA Pai Endowment Chair to MP at Manipal University, India, and the Manipal McGill Centre for Infectious Diseases (MAC ID). MP is also supported by a Canada Research Chair award. SKH has a fellowship from the FRQS.

Author Contributions

MN and SH conceptualized the article, wrote the protocol, and screened and extracted the papers. SB resolved disagreements in inclusion and extraction. SH performed the analysis. MN and SH drafted the manuscript. SB, KS, and MP critically reviewed the protocol and manuscript.
References

1. UNAIDS. 2014. The Gap Report.
2. Cepheid Technical Support. Xpert® HIV-1 Viral Load Training. http://www.cepheid.com/en/cepheid-solutions-uk/clinical-ivd-tests/virology/xpert-hiv-1-viral-load Accessed December 11, 2016.
3. World Health Organization. 2017. WHO monitoring of Xpert MTB/RIF roll-out. URL: http://www.who.int/tb/areas-of-work/laboratory/mtb-rif-rollout/en/ (accessed on 9 Feb 2017). Accessed
4. Cazabon D, Suresh A, Oghor C, Qin ZZ, Kik SV, Denkinger CM, Pai M. 2017. Implementation of Xpert MTB/RIF in 22 high tuberculosis burden countries: are we making progress? Eur Respir J 50.
5. World Health Organization. 2016. Global Tuberculosis Report. World Health Organization.
6. Roberts T, Cohn J, Bonner K, Hargreaves S. 2016. Scale-up of Routine Viral Load Testing in Resource-Poor Settings: Current and Future Implementation Challenges. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America 62:1043-1048.
7. World Health Organization. 2017. WHO Prequalification of In Vitro Diagnostics PUBLIC REPORT.
8. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Moher D. 2009. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med 6:e1000100.
9. Rucker G, Schwarzer G, Carpenter JR, Schumacher M. 2008. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 8:79.
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529-536.
11. Ceffa S, Luhanga R, Andreotti M, Brambilla D, Erba F, Jere H, Mancinelli S, Giuliano M, Palombi L, Marazzi MC. 2016. Comparison of the Cepheid GeneXpert and Abbott M2000 HIV-1 real time molecular assays for monitoring HIV-1 viral load and detecting HIV-1 infection. J Virol Methods 229:35-39.
12. Gous N, Scott L, Berrie L, Stevens W. 2016. Options to Expand HIV Viral Load Testing in South Africa: Evaluation of the GeneXpert® HIV-1 Viral Load Assay. PLOS ONE 11:e0168244.
13. DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Control Clin Trials 7:177-188.
14. Garrett NJ, Drain PK, Werner L, Samsunder N, Abdool Karim SS. 2016. Diagnostic accuracy of the point-of-care xpert HIV-1 viral load assay in a south african HIV clinic. J Acquir Immune Defic Syndr 72:e45-e48.
15. Gueudin M, Baron A, Alessandri-Gradt E, Lemee V, Mourez T, Etienne M, Plantier JC. 2016. Performance evaluation of the new HIV-1 quantification assay, Xpert HIV-1 viral load, on a wide panel of HIV-1 variants. J Acquir Immune Defic Syndr 72:521-526.
16. Jordan JA, Plantier JC, Templeton K, Wu AHB. 2016. Multi-site clinical evaluation of the Xpert HIV-1 viral load assay. J Clin Virol 80:27-32.

17. Mor O, Gozlan Y, Wax M, Mileguir F, Rakovsky A, Noy B, Mendelson E, Levy I. 2015. Evaluation of the RealTime HIV-1, Xpert HIV-1, and Aptima HIV-1 Quant Dx Assays in Comparison to the NucliSens EasyQ HIV-1 v2.0 Assay for Quantification of HIV-1 Viral Load. J Clin Microbiol 53:3458-3465.

18. Moyo S, Mohammed T, Wirth KE, Prague M, Bennett K, Holme MP, Mupfumi L, Sebogodi P, Moraka NO, Boleo C, Maphorisa CN, Seraise B, Gaseitsiwe S, Musonda RM, van Widenfelt E, Powis KM, Gaolathe T, Tchetgen Tchetgen EJ, Makhema JM, Essex M, Lockman S, Novitsky V. 2016. Point-of-care Cepheid Xpert HIV-1 Viral Load Test in Rural African Communities is Feasible and Reliable. J Clin Microbiol doi:http://dx.doi.org/10.1128/JCM.01594-16:12.

19. Bruzzzone B, Caligiuri P, Nigro N, Arcuri C, Delucis S, Di Biagio A, Icardi G. 2017. Xpert HIV-1 viral load assay and versant HIV-1 RNA 1.5 assay: A performance comparison. Journal of Acquired Immune Deficiency Syndromes 74:e86-e88.

20. Nash M, Ramapuram J, Kaiya R, Huddart S, Pai M, Baliga S. 2017. Use of the GeneXpert tuberculosis system for HIV viral load testing in India. Lancet Glob Health 5:e754-e755.

21. Avidor B, Matus N, Girshengorn S, Achsanov S, Gielman S, Zeldis I, Schweitzer I, Adler A, Turner D. 2017. Comparison between Roche and Xpert in HIV-1 RNA quantitation: a high concordance between the two techniques except for a CRF02_AG subtype variant with high viral load titters detected by Roche but undetected by Xpert. Journal of Clinical Virology 93:15-19.

22. Kulkarni S, Jadhav S, Khopkar P, Sane S, Londhe R, Chimanpure V, Dhilpe V, Ghate M, Yelagare T, Panchal N, Rahane G, Kadam D, Gaikwad N, Rewari B, Gangakhedkar R. 2017. GeneXpert HIV-1 quant assay, a new tool for scale up of viral load monitoring in the success of ART programme in India. BMC Infectious Diseases 17 (1) (no pagination).

23. Swathirajan CR, Vignesh R, Boobalan J, Solomon SS, Saravanan S, Balakrishnan P. 2017. Performance of point-of-care Xpert HIV-1 plasma viral load assay at a tertiary HIV care centre in Southern India. J Med Microbiol doi:10.1099/jmm.0.00514.

24. Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20:W13-W23.
Figure 1. PRISMA Flow Chart.

Figure 2. Forest plot for Pearson and Spearman correlation coefficients from comparison of VL values by Xpert and a reference test for VL.

Figure 3. Forest plot for Bland-Altman (BA) correlation coefficients from comparison of VL values by Xpert and a reference test for VL.
| Author          | Year | # enrolled | # quantifiable on index and reference test | Location       | Patient Population                  | Age | % Female | % on ART | Reference          |
|-----------------|------|------------|------------------------------------------|----------------|-------------------------------------|-----|----------|----------|---------------------|
| Avidor (21)     | 2017 | 383        | 254                                      | Israel         | HIV+, routine care                  |     |          |          | Roche Taqman       |
| Bruzzone (19)   | 2017 | 50         | 41                                       | Italy          | HIV+ lab samples                    |     |          |          | VERSANT             |
| Celfa (11)      | 2015 | 300        | 274                                      | Malawi         | Pediatric and adult HIV+, routine care |     |          |          | Abbott              |
| Garrett (14)    | 2016 | 42         | 42                                       | South Africa   | HIV+ women in CAPRISA 002 study     | 33  | 100      | 66       | Roche Taqman       |
| Gous, lab samples (12) | 2016 | 42       | 20                                       | South Africa   | Quality assessment HIV+ lab samples |     |          |          | Roche Taqman       |
| Gous, patient samples (12) | 2016 | 158     | 53                                       | South Africa   | HIV+, routine care                  | 42  | 54       | 100      | Roche Taqman       |
| Gouadin (15)    | 2016 | 295        | 162                                      | France         | HIV+, routine care                  |     |          |          | Abbott              |
| Jordan (16)     | 2016 | 764        | 390                                      | Europe, US     | Adult HIV+, routine care            | 45  | 28       |          | Abbott              |
| Kulkarni (22)   | 2017 | 219        | 167                                      | India          | HIV+, routine care                  | 38  | 52       |          | Abbott              |
| Mor (17)        | 2015 | 404        | 146                                      | Israel         | HIV+, routine care                  |     |          |          | NucliSens v2.0      |
| Moyo (18)       | 2016 | 302        | 277                                      | Botswana       | HIV+, ART-naïve, rural patients     |     |          |          | Abbott              |
| Nash (20)       | 2017 | 246        | 89                                       | India          | Adult HIV+, routine care            | 41  | 40       | 70       | Roche Taqman       |
| Swathirajan (23)| 2017 | 103        | 96                                       | India          | HIV+, routine care                  |     |          |          | Abbott              |

Full names and manufacturers of reference tests are as follows: Abbott = Abbott RealTime HIV-1 m2000rt, Abbott Molecular. Roche Taqman = COBAS® AmpliPrep / COBAS® TaqMan®, Roche Diagnostics. Versant = NucliSens EasyQ® HIV-1 v2.0, bioMérieux, VERSANT HIV-1 RNA 1.5 Assay, Siemens HealthCare Diagnostics.
TABLE 2: Study Quality Assessment Using Modified QUADAS2

| Study          | Risk of selection bias | Risk of non-generalizability | Risk of flow bias |
|----------------|------------------------|------------------------------|------------------|
| Avidor, 2017   | Unclear                | Unclear                      | Low              |
| Bruzzone, 2017 | Unclear                | Low                          | Low              |
| Cefal, 2015    | Low                    | Low                          | Low              |
| Garrett, 2016  | Low                    | Low                          | Low              |
| Gous, 2016     | Unclear                | Unclear                      | Unclear          |
| Gous, 2016     | Unclear                | Low                          | Unclear          |
| Gueudin, 2016  | Unclear                | Low                          | Low              |
| Jordan, 2016   | Low                    | Low                          | High             |
| Kulkarni, 2017 | Unclear                | Unclear                      | Low              |
| Mor, 2015      | Unclear                | Unclear                      | Low              |
| Moyo, 2016     | Unclear                | Unclear                      | Unclear          |
| Nash, 2017     | Low                    | Low                          | Low              |
| Swathirajan, 2017 | Unclear              | Unclear                      | Low              |

1 Risk of bases assessed using modified QUADAS2 (10). “Unclear” risk denotes there was insufficient information provided in the paper to assess the particular bias.
TABLE 3: Diversity of measures of agreement reported in Xpert HIV evaluations

| Measure                              | Avidor (2017) | Bruzzone (2017) | Ceffa (2015) | Garrett (2016) | Gous, lab samples (2016) | Gous, patient samples (2016) | Gueudin (2016) | Jordan (2016) | Kulkarni (2017) | Mor (2015) | Moyo (2016) | Nash (2017) | Swathirajan (2017) |
|--------------------------------------|---------------|-----------------|--------------|----------------|--------------------------|---------------------------|----------------|---------------|----------------|-------------|--------------|------------|-------------------|
| Pearson correlation                  |               |                 |              |                | **                       | **                        |                | **            |                |             |              |            |                   |
| Spearman correlation                 | **            | **              | **           | **             | **                       | **                        | **              |               | **            | **           | **           | **         |                   |
| Bland-Altman                         | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Passing Bablok Regression            | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Deming Regression                    | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Percent agreement, binary threshold  | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Percent agreement, 3+ categories     | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Percent Similarity                   | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Percent Similarity coefficient of variation | **         | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Kappa Statistic                      | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |
| Concordance Correlation              | **            | **              | **           | **             | **                       | **                        | **              | **            | **            | **           | **           | **         |                   |

*Study estimated Pearson correlation as a secondary metric so the Spearman correlation was used for this analysis.

**Study performed a Bland-Altman analysis but did not report a measure of variance and thus was excluded from the relevant analyses.

***Study provided a Bland-Altman plot but did not report the numerical mean log difference which we deemed necessary to be classified as a Bland-Altman analysis.
Records identified through database searching (n = 84)

Records after duplicates removed (n = 51)

Records screened (n = 51) → Records excluded (n = 29)

Full-text articles assessed for eligibility (n = 22)

Articles included (n = 12) Studies included (n = 13)

Excluded (n = 10) Reason for exclusion: Conference or poster abstract (n = 10)

Studies included in quantitative synthesis (Crr. n = 11, BA n = 11)
| Study                  | Total | Correlation | COR   | 95%–CI       | Weight |
|------------------------|-------|-------------|-------|--------------|--------|
| **Pearson**            |       |             |       |              |        |
| Avidor (2017)          | 254   |             | 0.97  | [0.96; 0.98] | 12.8%  |
| Bruzzone (2017)        | 41    |             | 0.93  | [0.87; 0.96] | 11.5%  |
| Ceefa (2015)           | 274   |             | 0.95  | [0.94; 0.96] | 12.8%  |
| Jordan (2016)          | 390   |             | 0.98  | [0.98; 0.99] | 12.9%  |
| Kulkarni (2017)        | 167   |             | 0.89  | [0.85; 0.91] | 12.7%  |
| Mor (2015)             | 146   |             | 0.90  | [0.86; 0.93] | 12.6%  |
| Nash (2017)            | 89    |             | 0.96  | [0.94; 0.97] | 12.3%  |
| Swathirajan (2017)     | 96    |             | 0.81  | [0.73; 0.87] | 12.4%  |
| **Random effects model** | 1457 |             | 0.94  | [0.89; 0.97] | 100.0% |
| **Spearman**           |       |             |       |              |        |
| Garrett (2016)         | 42    |             | 0.94  | [0.89; 0.97] | 31.1%  |
| Gueudin (2016)         | 162   |             | 0.98  | [0.98; 0.99] | 34.2%  |
| Moyo (2016)            | 277   |             | 0.92  | [0.90; 0.94] | 34.7%  |
| **Random effects model** | 481  |             | 0.96  | [0.86; 0.99] | 100.0% |
