# Improved HIV-1 Viral Load Monitoring Capacity Using Pooled Testing With Marker-Assisted Deconvolution

*Tao Liu, PhD,* Joseph W. Hogan, ScD,*† Michael J. Daniels, ScD,‡§ Mia Coetzer, PhD,‖ Yizhen Xu, MS,* Gerald Bove, PhD,* Allison K. DeLong, MS,* Lauren Ledingham, BA,‖ Millicent Orido, BS,† Lameck Diero, MD,†¶ and Rami Kantor, MD‖*

**Objective:** Improve pooled viral load (VL) testing to increase HIV treatment monitoring capacity, particularly relevant for resource-limited settings.

**Design:** We developed marker-assisted mini-pooling with algorithm (mMPA), a new VL pooling deconvolution strategy that uses information from low-cost, routinely collected clinical markers to determine an efficient order of *sequential* individual VL testing and dictates when the sequential testing can be stopped.

**Methods:** We simulated the use of pooled testing to ascertain virological failure status on 918 participants from 3 studies conducted at the Academic Model Providing Access to Healthcare in Eldoret, Kenya, and estimated the number of assays needed when using mMPA and other pooling methods. We also evaluated the impact of practical factors, such as specific markers used, prevalence of virological failure, pool size, VL measurement error, and assay detection cutoffs on mMPA, other pooling methods, and single testing.

**Results:** Using CD4 count as a marker to assist deconvolution, mMPA significantly reduces the number of VL assays by 52% [confidence interval (CI): 48% to 57%], 40% (CI: 38% to 42%), and 19% (CI: 15% to 22%) compared with individual testing, simple mini-pooling, and mini-pooling with algorithm, respectively. mMPA has higher sensitivity and negative/positive predictive values than mini-pooling with algorithm, and comparable high specificity. Further improvement is achieved with additional clinical markers, such as age and time on therapy, with or without CD4 values. mMPA performance depends on prevalence of virological failure and pool size but is insensitive to VL measurement error and VL assay detection cutoffs.

**Conclusions:** mMPA can substantially increase the capacity of VL monitoring.

**Key Words:** antiretroviral monitoring, viral load pooling, viral load, virological failure

(*J Acquir Immune Defic Syndr* 2017;75:580–587)

## INTRODUCTION

Monitoring and maintaining viral load (VL) suppression in HIV-infected individuals on antiretroviral therapy (ART) is crucial to control the spread of HIV and to reduce HIV-related morbidity and mortality.[1–5] In resource-limited settings (RLS), diagnoses of ART failure based on clinical and CD4 count criteria result in high misclassification rates,[6–12] leading to World Health Organization (WHO) recommendation of routine VL testing for ART monitoring where available.[13] Despite this recommendation, VL testing in many RLS is still constrained by cost, technology, and capacity, and the feasibility of routine VL testing varies widely.[14,15] Although it is expected that more HIV care programs will establish or increase capacity for VL monitoring with time, and point of care VL testing and associated price reductions are emerging,[16] VL testing will most likely remain limited and expensive in RLS for years to come, particularly with the expected continued global increase in ART access.[17] Thus, improved VL testing capacity combining efficiency, diagnostic accuracy, and high clinical utility is important to increase access to this clinically essential diagnostic test.

In settings where VL monitoring is limited, the VL test is typically used as confirmation for treatment failure based on CD4 or clinical criteria.[18] We have developed methods for targeted triage of VL testing,[19] which was subsequently

Received for publication December 2, 2016; accepted April 12, 2017.

From the *Department of Biostatistics and Center for Statistical Sciences, Brown University School of Public Health, Providence, RI; †Academic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya; Departments of ‡Statistics and Data Sciences; §Integrative Biology, University of Texas at Austin, Austin, TX; ‖Division of Infectious Diseases, the Alpert Medical School of Brown University, Providence, RI; and ¶Department of Medicine, School of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya.

Supported by the National Institutes of Health Grants [R01-AI-108441] and [P30-AI-042853] (Providence-Boston Center for AIDS Research).

Preliminary results of this work have been presented at the 2015 Conference on Retroviruses and Opportunistic Infections; "Improved viral load monitoring capacity with rank-based algorithms for pooled assays"; February 23–26, 2015; Seattle, WA.

The authors have no conflicts of interest to disclose.

T.L., J.W.H., R.K., and M.J.D. contributed to the study concept and design. Data were provided by R.K., M.O., and L.D. and prepared by A.K.D., G. B., and Y.X. The analysis was done by T.L. and Y.X. T.L., J.W.H., M.J. D., R.K., M.C., Y.X., G.B., A.K.D., and L.L. participated in interpretation of study findings. All authors provided revision and approval of the final manuscript.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Correspondence to: Tao Liu, PhD, Department of Biostatistics and Center for Statistical Sciences, Brown University, Box GS 121-7, Providence, RI 02912 (e-mail: tliu@stat.brown.edu).

applied in a South African cohort.[20] Another approach to optimizing utilization of limited VL assays is pooled testing, which has long been applied to *qualitative* assays that yield binary results (eg, positive versus negative) for ascertaining individual disease status or estimating disease prevalence,[21–26] including HIV infection.[27–29] Pooled testing also can be used on *quantitative* assays such as VL that yield a measured value. The simplest pooling strategy involves a single test on a pooled sample that combines equal amounts of samples from several subjects, which determines whether further testing of all individual samples (ie, deconvolution of the pool) is needed. This basic approach is often referred as mini-pooling (MP), to distinguish it from more complex pooling strategies, such as hierarchical[22,30] and matrix pooling.[31,32] May et al[33] improved the MP strategy by testing the individual samples *sequentially* at the deconvolution stage using an algorithm to decide when sequential testing can be stopped. This method is referred as "mini-pooling with algorithm" (MPA).[14,31,34–38]

If the rank-ordering of VL values in a pool was known, the number of VL tests needed to resolve that pool would be minimized. We hypothesize that information on low-cost, routinely collected clinical markers (RCMs) that are correlated with VL, such as CD4 count, CD4%, ART adherence, treatment history, and time on ART, can be used to rank samples in a pool and thereby reduce the number of assays needed for deconvolution. For VL testing, which uses quantitative assays, we develop a pooling strategy called "marker-assisted mini-pooling with algorithm" (mMPA) that uses a predicted VL based on RCMs to rank-order individual samples. By testing samples with a higher risk of virological failure first, we show that mMPA requires, on average, fewer VL assays for deconvolution than MPA, without sacrificing diagnostic accuracy. Applied broadly, this approach has potential to substantially increase the capacity of comprehensive VL monitoring. The proposed algorithm is a method of informative retesting, first proposed by Bilder et al for qualitative assays.[22,23,39] Use of a quantitative assay introduces a modification of the deconvolution algorithm and issues such as assay measurement error and lower limit of detection (LLD), which must be incorporated into assessments of algorithm performance.

## METHODS

### Definitions of Virological Failure and Pool Positivity

At the individual level, virological failure is defined as VL exceeding a threshold value $C$. In RLS, the WHO recommends using $C = 1000$ copies/mL (for brevity, units are omitted henceforth). When $K$ samples are combined for a pooled test, pool positivity is defined using a threshold of $C_{\text{pool}} = C/K$ to account for the dilution effect. For example, for a pool of size $K = 3$, $C_{\text{pool}} = 1000/3 = 333$. If the pool VL is $\leq 333$, one can deduce that no individual sample can have VL $> 1000$, and no further tests are needed. However if a pool has VL $> 333$, one or more individuals may have VL $> 1000$, prompting the need for individual tests. See the study by May et al[33] for more examples.

### mMPA Procedure

The mMPA strategy uses information from RCMs that are correlated with individual VL to develop a risk score $S$ for rank-ordering individual samples. The score is expressed as an explicit function of the RCMs. Two examples include the value of a single marker, such as concurrent CD4 count, and predicted VL derived from a statistical model using multiple markers as inputs. The latter can be developed by fitting a prediction model on a representative sample of individuals where both the RCMs and the VL have been measured.

Given the risk scores, if sequential individual testing is indicated based on the pool VL, then mMPA assays the individual samples sequentially in the decreasing rank-order of their estimated risks. Thus, samples with higher risk of virological failure are tested first. Although ordering by $S$ cannot guarantee that the individual sample with the highest VL is tested first, correlation between $S$ and VL makes this more likely, and higher correlations imply higher likelihood of correct ordering. A schematic overview of mMPA is given by a numerical example in Figure 1. In general, after testing each individual sample, the average VL value among the remaining samples can be deduced. The remaining (say $K'$) samples can be regarded as a "sub-pool." Thus, if the sub-pool VL is less than $C/K'$, the sub-pool is negative and no more individual tests are needed; otherwise, we continue testing the next individual sample. So intuitively by rank-ordering and testing high-risk individuals first, the virological status of each individual can be quickly ascertained.

### Data Sources

Our empirical comparison of mMPA with other testing strategies is based on simulating the application of each VL testing strategy in a combined cohort of patients enrolled in 3 studies at the Academic Model Providing Access to Healthcare (AMPATH) in Eldoret, Kenya: the "Crisis" study (n = 191),[40] conducted in 2009–2011 to investigate the impact of the 2007–2008 postelection violence in Kenya on ART failure and drug resistance; the "second-line" study (n = 394),[41] conducted in 2011–2012 to investigate ART failure and drug resistance among patients on second-line ART; and the "TDF" (tenofovir) study (n = 333),[42] conducted in 2012–2013 to investigate the impact of WHO guidelines changes to recommend TDF-based first-line ART, on ART failure and drug resistance, see Supplemental Digital Content for further details, http://links.lww.com/QAI/B16. The individual studies and this combined analysis were reviewed by the Lifespan and AMPATH ethics review committees.

In each study, demographic, clinical, laboratory, and VL data were collected for all participants. We use these data to calculate 4 candidate risk scores: (1) concurrent CD4 count; (2) predicted VL value generated by fitting a generalized additive model (GAM)[43] of log-10 VL to gender, age, CD4 count, CD4%, time on current ART, and WHO staging; (3) predicted VL from fitting a GAM to the same predictors as in (2), but without CD4 count and CD4%; and (4) predicted VL generated from a random forests model[44] using the same predictors as (2). Scenario (3) is relevant for settings where CD4 counts are limited or not routinely collected,[45] which is

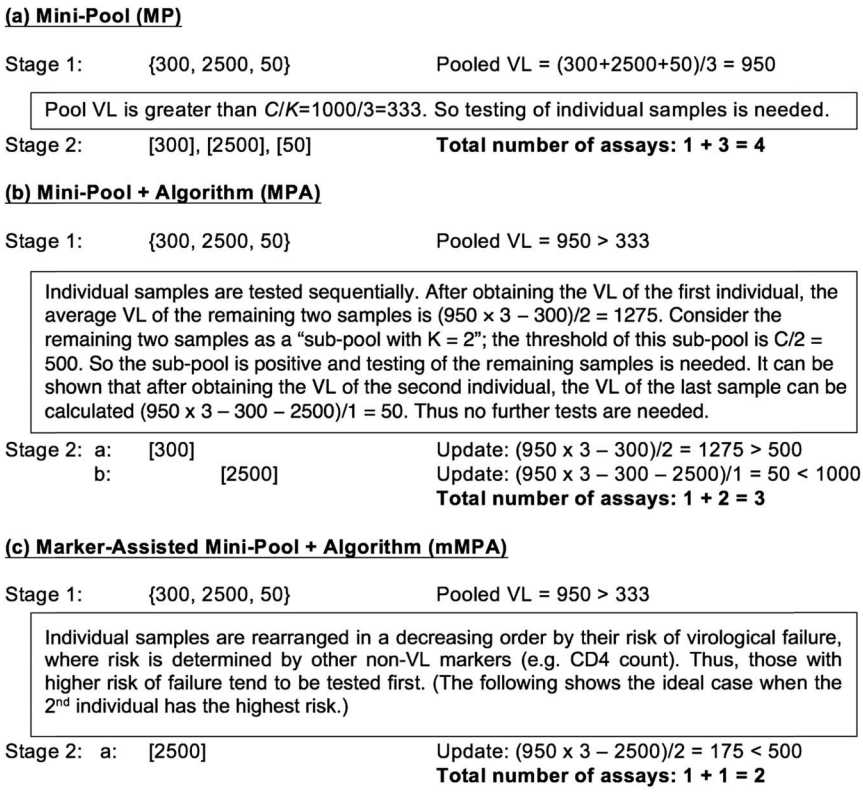


**FIGURE 1.** Schematic overview of MP, MPA, and mMPA using a numerical example of pooling $K = 3$ individual samples with VLs of 300, 2500, and 50 copies/mL, respectively. With a threshold of $C = 1000$, only the subject with VL = 2500 copies/mL has virological failure. Deconvolution is based on the 3 methods used in this article: (a) MP, (b) MPA, and (c) mMPA. Each method incorporates 2 stages, the first incorporating VL testing of the pooled sample and the second incorporating deconvolution, outlined by text boxes. In the first stage, common for all 3 methods, the pool is "positive" because its VL of 950 copies/mL is greater than the pool threshold $(C/K) = 333$, prompting the second stage. Using MP (a), all 3 individuals are then tested, for a total of 4 VL assays. Using MPA (b), sequential random testing of individual samples and a deconvolution algorithm determine the need for additional testing. In this example a total of 3 VL assays are needed. Using mMPA (c), with a risk score indicating that the second subject has the highest likelihood of VL failure and is thus tested first, only 2 VL assays are needed. {} denotes testing on the pool, and [] testing on an individual sample.

the case for some programs that have adopted yearly monitoring of VL.

## Simulation Study to Compare Testing Strategies

For each HIV testing strategy, we calculate the average number of tests required (ATR) to establish individual-level virological failure status per 100 individuals. The ATR for direct individual VL testing is 100; pooling methods will generally have ATR$<$100, with lower values indicating greater efficiency. Generally, the number of VL tests needed by a pooling algorithm for a fixed sample depends on the order in which individuals are placed into pools. Therefore, to obtain ATR for a single sample, we implement the pooling algorithm on 200 random permutations of the sample and calculated the average number of VL tests needed over the 200 permutations. Using permutations, we also calculate the sensitivity, specificity, and negative and positive predictive values (NPV and PPV, respectively) associated with each VL testing strategy.

For each pool formed by $K$ individuals, the pool VL is simulated using the average value of the individual VLs. For individuals with undetectable VL (n = 584), we impute values as follows: for those with VL $<$ 40 (n = 557), we impute a uniform random value between 0 and 40; next, for those with VL $<$ 400 (n = 27), we impute a value by sampling a random VL value from those with VL $<$ 400 (to avoid dependence on distributional assumptions).

To study robustness of each monitoring method, our evaluation further assesses the impact of the following factors on ATR and diagnostic accuracy: choice of risk score used by mMPA, prevalence of virological failure, pool size, assay measurement error, and assay LLD.

### Impact of Risk Score on ATR

Risk scores having a higher correlation with VL are expected to lead to higher testing efficiency for mMPA. Efficiency gain depends on the markers used to calculate the risk score and the method used to translate marker information into a score. The composite risk scores (2)–(4) (as described above) are generated using 2 types of models: GAM,[43] which allows nonlinear predictive effects of continuous markers; and random forests,[44] a tree-based approach machine learning method that allows interactions and nonlinear predictor effects.

Having the actual VL measurements also allows us to calculate the theoretical best possible ATR for mMPA using ranks of true VL values as a risk score. This "oracle" risk score is also included in our comparisons to provide a benchmark for the deconvolution algorithms.

Because mMPA uses risk scores to rank-order individual samples for sequential testing, we use rank (Spearman) correlation to quantify the relative strength of each risk score.

### Impact of Virological Failure Prevalence on ATR

Testing efficiency of mMPA depends on the prevalence of virological failure in the population. Higher prevalence implies higher proportion of positive pools, which generally reduces the advantage of all pooling strategies (including

mMPA). In addition to using the WHO-recommended threshold $C = 1000$, which corresponds to 16% failure prevalence in our sample, we further consider thresholds of $C = 500$, 1500, and 3000, yielding prevalences of 18%, 14%, and 11%, respectively.

### Impact of Pool Size on ATR

Pool size is another important parameter because larger pool sizes are technically more difficult to resolve and potentially more susceptible to measurement errors.[33] Moreover, pool size is a parameter for defining pool positivity (ie, $C/K$). In our study, we consider pool sizes ranging from $K = 3$ to 10.

### Impact of Measurement Error

Individual VL assays can be subject to measurement error, reducing diagnostic accuracy. To study its impact, we add random deviations to individual and pool VL values on the log scale. Deviations are generated using log-normal distributions with mean zero and SD σ. Three scenarios are considered: (1) $\sigma = 0$, or no measurement error; (2) $\sigma = 0.12$, applicable to commonly used assays[33]; and (3) $\sigma = 0.20$, likely representing an upper bound.[33] The impact of measurement error on the VL testing methods is quantified in sensitivity, specificity, PPV, and NPV.

### Impact of Lower Limit of Detection of VL Assays

VL assays used for treatment monitoring typically have LLD between 20 and 400.[13,46] The LLD impacts pooled testing in 2 ways. First, for a given virological failure threshold $C$, LLD dictates the maximum allowable pool size because of the definition of $C_{pool}$; for example, if $C = 1000$ and LLD = 100, then the allowable pool size is $K \leq 10$. Second, LLD impacts the second-stage pool deconvolution because the deconvolution algorithm of mMPA involves a sequence of subtractions of individual VLs from $K$ times the pool VL to determine whether tests of remaining samples are needed (Fig. 1). When an individual sample has a VL<LLD, the true VL is only known to be between 0 and LLD. In this case, it can be shown (see Supplemental Digital Content, http://links.lww.com/QAI/B16) that a conservative approach is to treat individuals below LLD as having VL = 0 and continue individual testing. Using this approach, assays with higher LLD will, on average, increase the number of tests needed by MPA and mMPA.

## RESULTS

### Participant Characteristics

Data were available from 918 participants, whose characteristics are summarized in Table 1. Overall, the demographic, clinical, and laboratory characteristics were similar between the 3 studies, with the exceptions of types of ART and prevalence of VL >1000. Table 2 summarizes the ATR for individual testing, MP, MPA, and mMPA for pool size $K = 5$, assuming LLD = 0 and no measurement error (scenarios using other values of LLD and measurement error are addressed below). We implemented mMPA using each of the 4 versions of the risk score as described in Methods.

### Deconvolution of VL Pooling With mMPA and Risk Score Impact

When applied to this combined cohort, mMPA requires between 46 and 50 assays to diagnose 100 individuals, depending on the risk score being used (Table 2). The most efficient risk score is derived using GAM with gender, age, CD4 count, CD4%, time on current ART, and WHO staging as predictors [ATR = 46, 95% confidence interval (CI): 41 to 50]; however, a GAM-based score without CD4 count still yields ATR = 50 (95% CI: 46 to 54). Compared to other pooling methods, mMPA reduces the ATR relatively by 19% (95% CI: 15% to 22%) compared to MPA, and by 40% (95% CI: 38% to 42%) compared to MP; all reductions are statistically significant as evidenced by the 95% confidence intervals excluding zero. Moreover, ATR for mMPA is relatively close to the best possible ATR of 37 based on the oracle risk score where correct ranks are known. The superior performance of mMPA has similar patterns when applied individually to the 3 studies (Supplemental Digital Content Table 1, http://links.lww.com/QAI/B16).

In the next sections, for simplicity, we used CD4 count as the risk score for mMPA.

### Impact of Virological Failure Prevalence and Pool Size

The prevalence of virological failure in the combined cohort was 16% based on the WHO recommended VL>1000 threshold. To study the effect of prevalence on ATR, we applied threshold values of 500, 1500, and 3000, yielding prevalences 11%, 14%, and 18%, respectively. We also varied the pool size from 3 to 10. Figure 2 depicts testing efficiency of mMPA with individual testing (IND), MP, and MPA in ATR under various failure prevalences and pool sizes, demonstrating the advantage of using mMPA in ATR. Both mMPA and MPA have lower ATR than IND and MP, especially when virological failure prevalence is low. MP may need more VL tests than IND when the prevalence is high and a large pool size is used (Fig. 2A). In general, mMPA is most advantageous when prevalence is low; for our data, when using a pool size of 5 or larger, mMPA has significantly the lowest ATR among all VL testing strategies.

Figure 2 also shows that the impact of pool size on pooled testing efficiency is nonlinear, and the optimal pool size that achieves the lowest ATR depends on the prevalence of virological failure. In general, pooling methods can achieve the best testing efficiency with a larger pool size when the prevalence is lower. This nonlinear pattern and low prevalence preferences are consistent with findings of other pooled testing procedures.[26,33] For this combined cohort, the optimal pool size for mMPA is 5 when the prevalence is 18%, 6 when the prevalence is 14%–16%, and 7 when the prevalence is 11%. mMPA maintains its superior performance for all pool sizes considered here.

### Impact of VL Assay Measurement Error

Figure 3 shows diagnostic accuracy as a function of virological failure prevalence for each testing method for

**TABLE 1.** Characteristics of Participants According to Study*

| | Crisis (n = 191) | Second Line (n = 394) | TDF (n = 333) | Total (n = 918) |
|---|---|---|---|---|
| Age, yrs | 42 (36–50) | 42 (36–49) | 41 (36–48) | 42 (36–49) |
| Gender, % | | | | |
| Female | 62 | 60 | 55 | 59 |
| Male | 38 | 40 | 45 | 41 |
| VL, copies/mL | 44 (<40 to <400) | <40 (<40 to 430) | <40 (<40 to <40) | <40 (<40 to 192) |
| VL > 1000, % | 15 | 21 | 10 | 16 |
| CD4 count (NA = 19), cells/μL | 361 (249–503) | 282 (182–419) | 336 (226–470) | 314 (210–453) |
| CD4% (NA = 19) | 23 (17–28) | 17 (12–23) | 21 (15–28) | 19 (14–26) |
| Most recent CD4 | 376 (262–553) | 273 (165–409) | 321 (210–467) | 311 (201–458) |
| Most recent CD4% | 20 (16–26) | 16 (11–21) | 2 (15–27) | 18 (13–25) |
| Chg of CD4/yr† | −14.3 (−121 to 94) | 0.5 (−75 to 170) | 0.0 (−63 to 112) | 0.0 (−85 to 127) |
| Chg of CD4%/yr† | 2.2 (0.0–5.3) | 2.0 (0.0–6.9) | 0.9 (−1.3 to 6.5) | 1.7 (0.0–6.5) |
| Yrs on current ART | 2.7 (2.1–4.4) | 1.1 (0.6–3.3) | 1.7 (1.2–2.2) | 1.8 (0.9–2.7) |
| Yrs on ART | 4.3 (3.6–5.1) | 5.9 (4.4–7.4) | 2.1 (1.4–4.6) | 4.5 (2.6–6.2) |
| ART adherence‡, % | | | | |
| Most | 0.0 | 0.8 | 0.3 | 0.4 |
| Some | 1.6 | 5.1 | 1.8 | 3.2 |
| None | 98.4 | 94.1 | 97.9 | 96.4 |
| WHO stage (NA = 4), % | | | | |
| Stage 1 | 18 | 14 | 24 | 19 |
| Stage 2 | 18 | 16 | 16 | 17 |
| Stage 3 | 46 | 48 | 43 | 45 |
| Stage 4 | 18 | 22 | 17 | 19 |
| ART, % | | | | |
| First line | 100 | 0 | 100 | 57 |
| Second line | 0 | 100 | 0 | 43 |

*Continuous variables are summarized by median (interquartile range) and categorical variables by percentages.
†Change = (value at enrollment − most recent before enrollment)/(time between the 2 visits).
‡ART adherence of last week by the self-report percentage: most (>50%), some (<50%), none (0%).
chg, change; NA, not available; TDF, tenofovir.

assay-level measurement errors SD = 0.12 on the log-10 scale. All methods have high specificity (>99%). Sensitivity of mMPA ranges between 80% and 90%, lower than IND and MP, but greater than MPA (<80%). NPV for mMPA exceeds 95% for all prevalences considered, better than MPA and only slightly lower than IND and MP. Moreover, mMPA has the highest PPV among all methods (although all are above 95%). mMPA outperforms MPA on all 4 accuracy measures because it requires fewer VL assays than MPA, which implies fewer subtractions during deconvolution and therefore fewer opportunities for measurement error to affect diagnostic decisions. Similar patterns in diagnostic accuracy are observed when the assay-level

**TABLE 2.** Efficiency Comparison Among Pooling Deconvolution Methods*

| Method | Risk Score | Spearman Correlation† | Average No. of VL Assays Required per 100 Subjects | Relative Reduction (%) in the No. of Assays Required | | |
|---|---|---|---|---|---|---|
| | | | | REF = IND | REF = MP | REF = MPA |
| IND | | | 100 | REF | | |
| MP | | | 80 (73–86) | −20 (−27 to −14) | REF | |
| MPA | | | 59 (54–63) | −41 (−46 to −37) | −26 (−28 to −24) | REF |
| mMPA | CD4 | 0.24 | 48 (43–52) | −52 (−57 to −48) | −40 (−42 to −38) | −19 (−22 to −15) |
| | GAM | 0.34 | 46 (41–50) | −54 (−59 to −50) | −43 (−45 to −41) | −23 (−25 to −19) |
| | GAM* | 0.25 | 50 (46–54) | −50 (−54 to −46) | −37 (−39 to −35) | −14 (−19 to −10) |
| | RF | 0.27 | 47 (43–52) | −53 (−57 to −48) | −41 (−43 to −38) | −20 (−23 to −16) |
| | ORS | 1.00 | 37 (34–39) | −63 (−66 to −61) | −54 (−55 to −53) | −38 (−39 to −36) |

*Pool size $K = 5$; numbers in parentheses are 95% confidence intervals obtained using the bootstrap method (with 500 resamples) where the intervals are the 2.5 and 97.5 percentiles of bootstrap distributions.
†Spearman (rank) correlation between risk score and VL.
GAM, generalized additive model; GAM*, GAM without CD4 markers as predictor; ORS, oracle rank score; REF, reference method; RF, random forest model.

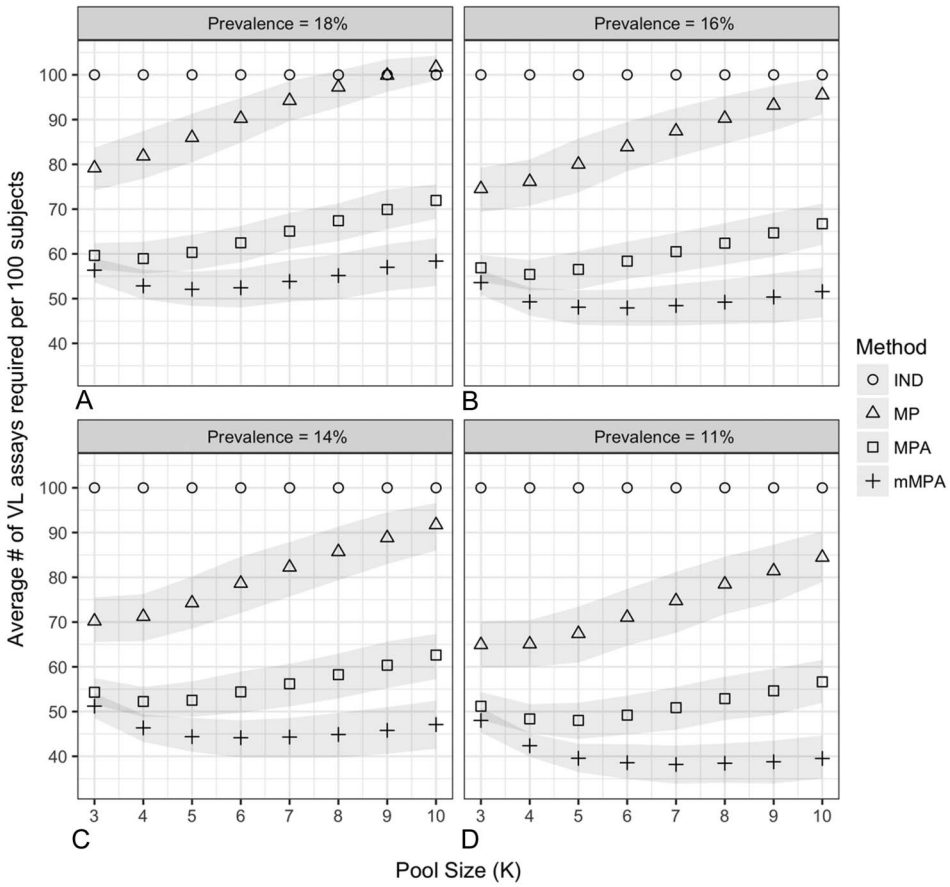


**FIGURE 2.** Impact of virological failure prevalence and pool size on pooling deconvolution methods. The figure demonstrates the impact of pool sizes K (*x* axis; from 3 to 10) and viral failure prevalence [(A) = 18%; (B) = 16%; (C) = 14%; and (D) = 11%] on the ATR (*y* axis) according to different testing strategies: IND (circles), MP (triangles), MPA (squares), and mMPA (plus signs). CD4 is used as a risk score for mMPA. The gray bounds indicate point-wise 95% confidence intervals obtained using the bootstrap method (with 500 resamples), where the intervals are the 2.5 and 97.5 percentiles of the bootstrap distributions.

measurement error has a SD of 0.20 (log-10 scale) (Supplemental Digital Content, Figure 1, http://links.lww.com/QAI/B16).

## Impact of LLD

Theoretically, a high LLD increases ATR for MPA and mMPA. In our data, the impact is negligible (Supplemental Digital Content, Figure 2, http://links.lww.com/QAI/B16) because the majority (557/918 = 61%) had VL $< 40$. Thus, treating those with VL below the LLD as VL = 0 had little impact on the deconvolution process and subsequent individual-level diagnoses resulting from pooled testing.

# DISCUSSIONS

mMPA is an informative retesting strategy to improve deconvolution of pooled HIV VL assays for ART monitoring. It uses an algorithm that relies on information from routinely collected clinical information to rank-order samples contributing to the pool, resulting in faster and more efficient deconvolution and fewer VL tests than other approaches. The efficiency gains realized by mMPA persist across variations in virological failure prevalence, pool size, assay measurement error, and LLD.

Bilder et al[22,23,39] developed single- and multi-stage informative retesting algorithms for pooled testing using qualitative assays and provided a comprehensive evaluation of their operating characteristics. The mMPA proposed here is designed for quantitative assays; the algorithm itself is similar to "informative Sterrett" algorithm[47] for qualitative assays, in which positive pools lead to an ordered sequence of tests on individual samples and on pools formed from remaining samples (quantitative assays do not require this latter step). However, there are important differences, such as a modified deconvolution algorithm using sequential subtractions, assay measurement error, and assay limit of detection, that are related to the implementation of mMPA.

In light of expanding VL monitoring in RLS, pooled VL testing methods, and MPA in particular, have demonstrated high potential to increase capacity of individual-level monitoring.[31,33,34,36–38,48] Our study demonstrates that by incorporating sequential ranked deconvolution, mMPA can further improve the efficiency of pooled VL testing. Analysis of data from western Kenya demonstrates reductions of 52% relative to individual testing and by 19% relative to MPA. In laboratory capacity and frequency of individual-level monitoring in programs with comparable characteristics, mMPA could double the number of virological failure ascertainments (a 108% increase in capacity) relative to individual testing or increase by 23% the capacity relative to MPA. Moreover, mMPA has greater sensitivity, PPV, and NPV than MPA and equally high specificity. Consequently, programmatic incorporation of mMPA would allow patients to get more, and perhaps more frequent (eg, twice yearly than the

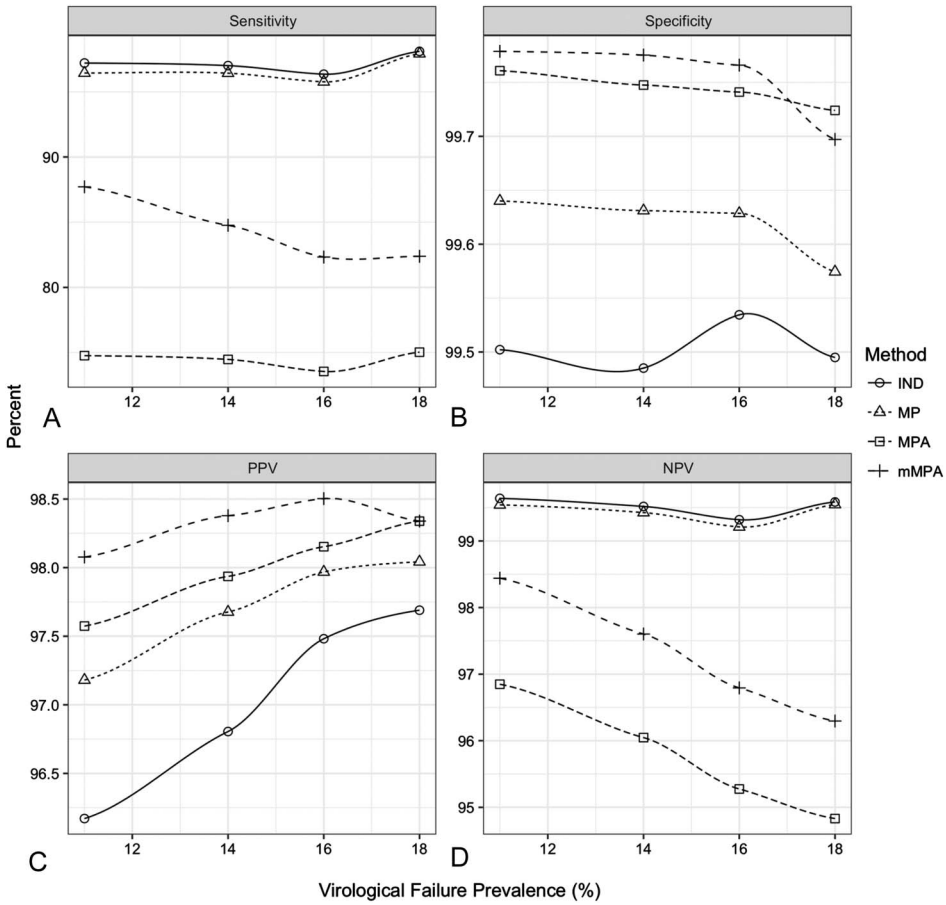


**FIGURE 3.** Impact of VL assay measurement error on diagnostic accuracy of pooling deconvolution methods. The figure demonstrates the impact of VL assay measurement errors on: IND (circles), MP (triangles), MPA (squares), and mMPA (plus signs). CD4 is used as a risk score for mMPA. The impact of measurement error on pooling deconvolution is quantified in (A) sensitivity, (B) specificity, (C) PPV, and (D) NPV. The measurement error is assumed to have a log-normal distribution with a zero mean and an SD of 0.12 on the log-10 scale.

recommended annual) VL monitoring. Such improvement in treatment monitoring capacity would provide more reliable guidance to physicians in clinical care and reduce the risk of maintaining patients on ineffective ART.

mMPA incorporates an estimated risk score using RCMs. Although we show the advantage of using composite risk scores from multiple markers, including one that does not use CD4 values, most of our analyses focused on CD4 count as the risk score for its simplicity. We show that CD4 count alone, although inadequate to determine virological status, still contains valuable information that could significantly improve the efficiency of pooled VL testing. Developing risk scores for risk assessment and prediction has been an important topic in areas such as statistics and machine learning. We are exploring the use of methods such as generalized boosting,[49] classification and regression trees,[50] and ensemble techniques such as super learner[51] to develop risk scores with even better diagnostic properties.

Some limitations of this study point to future research directions. First, our analysis relies on a simulated application of the methods to patients participating in 3 separate research studies, as VL data from routine clinical care in RLS are still limited. This data aggregation provides a large sample size and heterogeneous first- and second-line ART exposure for evaluating different testing strategies (our focus), but the combined cohort is not necessarily a representative sample of all patients in care. Degree of improved capacity depends on correlation between risk score and VL, which could differ in a randomly selected patient population. Second, factors such as time, cost, and practical constraints associated with mMPA testing, deconvolution, and implementation are not considered in this simulation study, yet are important in assessing the feasibility and benefit of this approach, particularly in RLS. These limitations have motivated a comparative evaluation of mMPA performance using actual patient samples, currently ongoing.

Despite some limitations, this study demonstrates the potential for mMPA to substantially improve capacity for VL monitoring relative to individual testing and other VL pooling strategies, a particularly important issue in light of urgent demand for expanding virological monitoring in RLS. To enable its application to specific settings, we have developed a software package called "mMPA" using R.[52] The package allows HIV care facilities to evaluate the capacity of ART monitoring for their patients when using pooled testing to meet their context (eg, virological failure prevalence, available clinical markers, pool size, and detection cutoff). More information about the package is provided in the Supplemental Digital Content, http://links.lww.com/QAI/B16.

## REFERENCES

1. Anderson AM, Bartlett JA. Changing antiretroviral therapy in the setting of virologic relapse: review of the current literature. *Curr HIV/AIDS Rep.* 2006;3:79–85.

2. Calmy A, Ford N, Hirschel B, et al. HIV viral load monitoring in resource-limited regions: optional or necessary? *Clin Infect Dis.* 2007;44: 128–134.
3. Vekemans M, John L, Colebunders R. When to switch for antiretroviral treatment failure in resource-limited settings? *AIDS*. 2007;21: 1205–1206.
4. Panel on Antiretroviral Guidelines for Adults and Adolescents. *Guidelines for the Use of Antiretroviral Agents in HIV-1-infected Adults and Adolescents*. Department of Health and Human Services. 2016. Available at: http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf. Accessed on 9/12/2016.
5. WHO. Technical and operational considerations for implementing HIV viral load testing. 2016. Available at: http://www.who.int/hiv/pub/arv/viral-load-testing-technical-update/en/. Accessed on 7/8/2016.
6. Bisson GP, Gross R, Bellamy S, et al. Pharmacy refill adherence compared with CD4 count changes for monitoring HIV-infected adults on antiretroviral therapy. *PLoS Med.* 2008;5:e109.
7. Mee P, Fielding KL, Charalambous S, et al. Evaluation of the WHO criteria for antiretroviral treatment failure among adults in South Africa. *AIDS.* 2008;22:1971–1977.
8. Keiser O, Macphail P, Boulle A, et al. Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy. *Trop Med Int Health.* 2009;14:1220–1225.
9. Meya D, Spacek LA, Tibenderana H, et al. Development and evaluation of a clinical algorithm to monitor patients on antiretrovirals in resource-limited settings using adherence, clinical and CD4 cell count criteria. *J Int AIDS Soc.* 2009;12:3.
10. Reynolds SJ, Nakigozi G, Newell K, et al. Failure of immunologic criteria to appropriately identify antiretroviral treatment failure in Uganda. *AIDS.* 2009;23:697–700.
11. Kiragga AN, Castelnuovo B, Kamya MR, et al. Regional differences in predictive accuracy of WHO immunologic failure criteria. *AIDS*. 2012; 26:768–770.
12. Kantor R, Diero L, DeLong A, et al. Misclassification of first-line antiretroviral treatment failure based on immunological monitoring of HIV infection in resource-limited settings. *Clin Infect Dis.* 2009;49: 454–462.
13. WHO. March 2014 supplement to the 2013 consolidated guideline on use of antiretroviral drugs for treating and preventing HIV infection. 2014. Available at: http://apps.who.int/iris/bitstream/10665/104264/1/9789241506830_eng.pdf?ua=1. Accessed on 8/11/2016.
14. Rowley CF. Developments in CD4 and viral load monitoring in resource-limited settings. *Clin Infect Dis.* 2014;58:407–412.
15. UNAIDS. *Speed Up Scale-up: Strategies, Tools and Policies to Get the Best HIV Treatment to More People Sooner*. Geneva, Switzerland: Médecins Sans Frontières; 2012:1–23. 78.
16. Stevens W, Gous N, Ford N, et al. Feasibility of HIV point-of-care tests for resource-limited settings: challenges and solutions. *BMC Med.* 2014;12:1–8.
17. Piot P, Abdool Karim SS, Hecht R, et al. Defeating AIDS—advancing global health. *The Lancet.* 2015;386:171–218.
18. Roberts T, Cohn J, Bonner K, et al. Scale-up of routine viral load testing in resource-poor settings: current and future implementation challenges. *Clin Infect Dis.* 2016;62:1043–1048.
19. Liu T, Hogan JW, Wang L, et al. Optimal allocation of gold standard testing under constrained availability: application to assessment of HIV treatment failure. *J Am Stat Assoc.* 2013;108:1173–1188.
20. Koller M, Fatti G, Chi BH, et al. Risk charts to guide targeted HIV-1 viral load monitoring of ART: development and validation in patients from resource-limited settings. *J Acquir Immune Defic Syndr.* 2015;70: e110–119.
21. Dorfman R. The detection of defective members of large populations. *Ann Math Stat.* 1943;14:436–440.
22. Bilder CR, Tebbs JM, Chen P. Informative retesting. *J Am Stat Assoc.* 2010;105:942–955.
23. McMahan CS, Tebbs JM, Bilder CR. Informative Dorfman screening. *Biometrics*. 2012;68:287–296.
24. Balding DJ, Bruno WJ, Torney DC, et al. *Genetic Mapping and DNA Sequencing: A Comparative Survey of Non-adaptive Pooling Designs*. New York, NY: Springer-Verlag; 1996.
25. Brookmeyer R. Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics*. 1999;55:608–612.
26. Kim HY, Hudgens MG, Dreyfuss JM, et al. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*. 2007;63:1152–1163.
27. Stekler J, Swenson PD, Wood RW, et al. Targeted screening for primary HIV infection through pooled HIV-RNA testing in men who have sex with men. *AIDS*. 2005;19:1323–1325.
28. Tu XM, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening. *Biometrika*. 1995;82:287–297.
29. Westreich DJ, Hudgens MG, Fiscus SA, et al. Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests. *J Clin Microbiol.* 2008;46:1785–1792.
30. Quinn TC, Brookmeyer R, Kline R, et al. Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS*. 2000;14:2751–2757.
31. van Zyl GU, Preiser W, Potschka S, et al. Pooling strategies to reduce the cost of HIV-1 RNA load monitoring in a resource-limited setting. *Clin Infect Dis.* 2011;52:264–270.
32. Litvak E, Tu XM, Pagano M. Screening for the presence of a disease by pooling sera samples. *J Am Stat Assoc.* 1994;89:424–434.
33. May S, Gamst A, Haubrich R, et al. Pooled nucleic acid testing to identify antiretroviral treatment failure during HIV infection. *J Acquir Immune Defic Syndr.* 2010;53:194–201.
34. Pannus P, Fajardo E, Metcalf C, et al. Pooled HIV-1 viral load testing using dried blood spots to reduce the cost of monitoring antiretroviral treatment in a resource-limited setting. *J Acquir Immune Defic Syndr.* 2013;64:134.
35. Smith DM, May SJ, Perez-Santiago J, et al. The use of pooled viral load testing to identify antiretroviral treatment failure. *AIDS*. 2009;23:2151.
36. Chohan BH, Tapia K, Merkel M, et al. Pooled HIV-1 RNA viral load testing for detection of antiretroviral treatment failure in Kenyan children. *J Acquir Immune Defic Syndr.* 2013;63:e87–e93.
37. Tilghman M, Tsai D, Buene TP, et al. Pooled nucleic acid testing to detect antiretroviral treatment failure in HIV-infected patients in Mozambique. *J Acquir Immune Defic Syndr.* 2015;70:256–261.
38. Tilghman MW, Guerena DD, Licea A, et al. Pooled nucleic acid testing to detect antiretroviral treatment failure in Mexico. *J Acquir Immune Defic Syndr.* 2011;56:e70.
39. Bilder CR, Tebbs JM. Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations. *Stat Med.* 2012; 31:3261–3268.
40. Mann M, Diero L, Kemboi E, et al. Antiretroviral treatment interruptions induced by the Kenyan postelection crisis are associated with virological failure. *J Acquir Immune Defic Syndr.* 2013;64:220–224.
41. Diero L, DeLong A, Schreier L, et al. High HIV resistance and mutation accrual at low viral loads upon 2nd-line failure in western Kenya. Paper presented at: Conference on Retroviruses and Opportunistic Infections; March 3–6, 2014; Boston, MA.
42. Brooks K, Diero L, DeLong A, et al. Treatment failure and drug resistance in HIV-positive patients on tenofovir-based first-line antiretroviral therapy in western Kenya. *J Int AIDS Soc.* 2016;19:20798.
43. Hastie T, Tibshirani R. Generalized additive models. *Stat Sci.* 1986;1: 297–318.
44. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell.* 1998;20:832–844.
45. Sax PE. Editorial Commentary: can we break the habit of routine CD4 monitoring in HIV care? *Clin Infect Dis.* 2013;56:1344–1346.
46. Swenson LC, Cobb B, Geretti AM, et al. Comparative performances of HIV-1 RNA load assays at low viral load levels: results of an international collaboration. *J Clin Microbiol.* 2014;52:517–523.
47. Sterrett A. On the detection of defective members of large populations. *Ann Math Stat.* 1957;28:1033–1036.
48. Kim SB, Kim HW, Kim H-S, et al. Pooled nucleic acid testing to identify antiretroviral treatment failure during HIV infection in Seoul, South Korea. *Scand J Infect Dis.* 2014;46:136–140.
49. Friedman J, Hastie T, Rosset S, et al. Discussion of boosting papers. *Ann Stat.* 2004;28:337–407.
50. Breiman L, Friedman J, Stone CJ, et al. *Classification and Regression Trees*. 1st ed. London, United Kingdom: Chapman and Hall/CRC; 1984.
51. Sinisi SE, Polley EC, Petersen ML, et al. Super learning: an application to the prediction of HIV-1 drug resistance. *Stat Appl Genet Mol Biol.* 2007;6:7.
52. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015.