2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Implications of Measurement Assay Type in Design of HIV Experiments

LaMont Cannon[1], Aditya Jagarapu[1], Cesar A. Vargas-Garcia[2], Michael J. Piovoso[2] and Ryan Zurakowski[1,2],

Abstract— **Time series measurements of circular viral episome (2-LTR) concentrations enable indirect quantification of persistent low-level Human Immunodeficiency Virus (HIV) replication in patients on Integrase-Inhibitor intensified Combined Antiretroviral Therapy (cART). In order to determine the magnitude of these low level infection events, blood has to be drawn from a patients at a frequency and volume that is strictly regulated by the Institutional Review Board (IRB). Once the blood is drawn, the 2-LTR concentration is determined by quantifying the amount of HIV DNA present in the sample via a PCR (Polymerase Chain Reaction) assay. Real time quantitative Polymerase Chain Reaction (qPCR) is a widely used method of performing PCR; however, a newer droplet digital Polymerase Chain Reaction (ddPCR) method has been shown to provide more accurate quantification of DNA. Using a validated model of HIV viral replication, this paper demonstrates the importance of considering DNA quantification assay type when optimizing experiment design conditions. Experiments are optimized using a Genetic Algorithm (GA) to locate a family of suboptimal sample schedules which yield the highest fitness. Fitness is defined as the expected information gained in the experiment, measured by the Kullback-Leibler Divergence (KLD) between the prior and posterior distributions of the model parameters. We compare the information content of the optimized schedules to uniform schedules as well as two clinical schedules implemented by researchers at UCSF and the University of Melbourne. This work shows that there is a significantly greater gain information in experiments using a ddPCR assay vs. a qPCR assay and that certain experiment design considerations should be taken when using either assay.**

I. INTRODUCTION

Advances in Combined Antiretroviral Therapy (cART) for treatment of the Human Immunodeficiency Virus (HIV) have drastically reduced AIDS related mortality rates worldwide [10]. Clinical analysis has shown that cART is able to suppress viral levels below the limit of detection; however, complete eradication is not achieved [4], [10], [11]. An important theory to explain the inability to eliminate the virus with cART is that there are low levels of on-going viral replication occurring in sanctuary cites where drug concentrations are depressed [3], [4], [19], [24]. Preliminary studies have been conducted to test for biomarkers of replication yielding inconclusive results [3], [19], [24]. This has given rise to a need to investigate new methods for designing clinical trials to test for on-going replication to achieve more conclusive results.

Previous studies have sought to detect on-going viral replication by intensifying antiretroviral therapy with an integrase inhibitor [3], [19]. In the presence of an integrase inhibitor, viral DNA is unable to integrate into the host genome. Host nuclear enzymes convert this un-integrated DNA into circles with two adjacent long-terminal repeat sequences. These converted DNA elements, which are referred to as 2-LTR circles, are an indicator of on-going replication [1], [3], [19]. In vivo 2-LTR concentrations are estimated using the polymerase chain reaction (PCR) method on blood samples drawn from patients under integrase inhibitor intensification. If a high level of on-going replication is present, a transient increase in the 2-LTR concentration is expected [1], [3], [19], [24]. There will initially be a sharp increase in production of 2-LTR circles as the new infections are inhibited but the production will then decrease since the success rate of infection events is drastically decreased [1], [3], [19], [21], [24].

Due to the expected dynamics in the presence of on-going replication, a dynamic model of 2-LTR concentration is able to quantify the amount of on going replication [1], [21]. However, do to Institutional Review Board (IRB) limits on the amount of blood that can be drawn during a clinical trial, as well as costs associated with acquiring patient samples, only a few samples can be taken to fit the model [1], [21], [22]. There is also inherent noise in the quantification of the HIV DNA via PCR assay. Real time qPCR assays, while inexpensive and popular, yield far noisier estimates of DNA concentration than ddPCR assays [20]. Taking into consideration the limited number of measurements and the accuracy of the measurement assay we investigate suboptimal designs of an experiment to quantify the level of ongoing replication. We then compare our optimized deign to designs in similar HIV 2-LTR experiments.

II. METHODS

A. HIV-1 2-LTR Model

In order to capture the dynamics of the 2-LTR circles following intensification with the integrase inhibitor, we use a two-state ordinary differential equation model previously developed by Luo et al. [1], [2]. The two states describe both the concentration of 2-LTR circles and the concentration of actively infected CD4+ T Cells in the blood. These equations

*This work was conducted at the University of Delaware Biomedical Engineering Department. Research reported in the publication was supported by the National Institutes of Health (NIH) grant AI110288, and Merck. The content is solely the responsibility of the authors and does not necessarily represent the official view of the funders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

[1]University of Delaware Biomedical Engineering Department, Newark, DE 19716, USA lccannon@udel.edu, srihari@udel.edu,

[2]University of Delaware Electrical and Computer Engineering Department Newark, DE 19716, USA cavargar@udel.edu, piovoso@udel.edu, ryanz@udel.edu

have the form:

$$\begin{aligned}\dot{\mathbf{y}} &= -(1-(1-\eta_{II}\mathbf{u_{II}})R)a\mathbf{y} + \mathbf{y_e} \\ \dot{\mathbf{c}} &= \phi k_{II}(1-\eta_{II}\mathbf{u_{II}})R)a\mathbf{y} + k_{II}\eta_{II}\mathbf{u_{II}}Ra\mathbf{y} - \delta\mathbf{c}\end{aligned} \tag{1}$$

where $\mathbf{y}$ is the concentration of actively infected cells and $\mathbf{c}$ is the concentration of 2-LTR circles. The term y_e is the entry rate into the blood of actively infected cells from exogenous sources unaffected by the integrase inhibitor, such as activation of latently infected cells. One assumption that is made is that the dynamics have reach steady state prior to treatment intensification with the integrase inhibitor. Based on this assumption, Luo et al. are able to simplify the equation for the 2-LTR concentration to:

$$\begin{aligned}\mathbf{c}(t) &= \mathbf{c}(\infty) + (\mathbf{c}(0) - \mathbf{c}(\infty))e^{-\delta t} \\ &+ \mathbf{c}(\infty)\frac{\delta\eta_{II}R(e^{-\delta t} - e^{-a(1-(1-\eta_{II})R)t})}{(1-R)(a(1-(1-\eta_{II})R) - \delta)}\end{aligned} \tag{2}$$

with a steady state initial value of

$$\mathbf{c}(0) = \frac{k_{II}\mathbf{y_e}\phi R}{\delta(1-R)} \tag{3}$$

and the final value of

$$\mathbf{c}(\infty) = \frac{k_{II}\mathbf{y_e}(\phi + \eta_{II} - \phi\eta_{II})R}{\delta(1-(1-\eta_{II})R)} \tag{4}$$

Parameter definitions and units are defined below in Table I.

TABLE I
MODEL PARAMETER DEFINITIONS

Parameter	Definition	Units
$\mathbf{y}$	concentration of infected cells	cells/10^6PBMC
$\mathbf{c}$	concentration of 2-LTR circles	2LTR/10^6PBMC
R	probability infected cell infects a target cell in a generation	unitless
a	death rate of actively infected cells	day^{-1}
$\mathbf{y}_e$	rate of exogenous production of infected	infected cells/ 10^6PBMCxDay
η_{II}	Ratio-reduction in R following integrase inhibitor intensification	unitless
$\mathbf{u}_{II}$	binary variable: 1 when integrase inhibitor is applied and 0 when it is not	unitless
ϕ	Ratio of probability of 2-LTR formation with integrase inhibitor vs without	unitless
k_{II}	The probability of 2-LTR circle formation when integrase inhibitor is present	2LTR/ infected cells
δ	decay rate of 2-LTR circles	day^{-1}

B. qPCR Measurement Noise

The majority of HIV studies have been conducted using quantitative PCR (qPCR) quantification methods. While qPCR often provides useful data, in many cases results are inconclusive due to the high level of noise associated with the assay [18], [20].

The first source of noise in a PCR assay is the noise due to sampling from the blood. Given a small sample size, the number of copies of HIV 2-LTR particles follows a Poisson distribution with a probability mass function (PMF) as follows.

$$P(n|v\times c) = \frac{(v\times c)\cdot e^{-(v\times c)}}{n!} \tag{5}$$

where v is the volume sampled, c is the particle concentration in the blood, and n is the total number of particles present in the sample.

The next step in the qPCR process is amplification of the DNA and quantification using a labeling probe such as the TaqMan probe [18], [20]. This results in a qPCR assay with the probability mass function

$$P(m|n) = \begin{cases} \frac{exp\left(-\frac{(ln\ m - ln\frac{n}{v})^2}{2\sigma(n)^2}\right)}{\sqrt{2\pi}m\sigma(n)}, & m > 0 \\ 1, & m = 0 \\ 0, & Otherwise \end{cases} \tag{6}$$

where $\sigma(n) = ln10 \times 10^{-0.21-0.24log_{10}n}$ is the equation for the log normal standard deviation of the qPCR growth process as a function of viral concentration

C. ddPCR Measurement Noise

Droplet Digital Polymerase Chain Reaction (ddPCR) quantification is very similar to the qPCR with one primary difference. After the sample is taken from the blood it is separated into thousands of nanoliter sized droplets prior to amplification and labeling. Using microfluidic technology the droplets are then classified as either positive for containing viral particles or negative for containing viral particles. This method is much more accurate because the probability that there are multiple copies per droplet is very low, therefore the total positive droplet concentration is able to able to provide a good estimate of the true sample concentration [20]. For the ddPCR assay the probability of d positive droplets given a blood concentration can be modeled as a simple binomial function

$$B(N,p) \tag{7}$$

where N is the total number of droplets and $p = 1 - e^{(\frac{c\times v_d}{N})}$. The concentration of viral particles int he blood is c and the droplet volume is represent by v_d.

A further analysis of these equations shows us that the noise in the qPCR measurement is largely due to the noise in the qPCR measurement process whereas the noise in the ddPCR measurement assay is largely dominated by the noise of the sampling process [18]. For this reason the ddPCR assay yields a much more accurate estimate of the true concentration than the qPCR assay [20].

D. Unscented Transform (UT) Sigma Point Patients

In order to evaluate a sampling schedule's performance across the prior distribution while maintaining computational tractability, we apply an unscented transform to our prior distribution to obtain 11 simulated sigma point patients that

approximately represent the distribution of possible patients [16], [21].

$$X_i = \begin{cases} \mu, & i = 0 \\ \mu + \sqrt{N\Sigma_i}, & 1 \leq i \leq N \\ \mu - \sqrt{N\Sigma_i}, & N < i \leq 2N \end{cases} \\ X_i = (A_i, \phi_i, R_i, \eta_{IIi}, \delta_i) \tag{8}$$

where each X_i is the separate set of parameters for each sigma point patient. The μ and Σ_i terms are the mean of the prior distribution and the i_{th} column of the covariance matrix of the prior distribution respectively. N is the total number of dimensions in our prior distribution, which in this case is a five dimensional distribution. KLD is calculated for each patient X_i and then averaged to find the expected KLD for each prospective sampling schedule.

E. Markov Chain Monte Carlo Methodology

For each candidate schedule we constructed simulated data for all 11 simulated sigma point patients based on our model and measurement noise. The posterior distributions for parameter set $\Theta_i(A_i, \phi_i, R_i, \eta_{IIi}, \delta_i)$ are constructed for each patient i using a Markov Chain Monte Carlo technique. We define $\mathbf{c}(t_k, \Theta_i)$ as the true concentration of 2-LTR DNA circles in the blood measured at sample point k for patient i using parameter set Θ_i. We assume measurement noise consistent with a droplet digital polymerase chain reaction (ddPCR) quantification assay [20], [21]. This leads to measurements as

$$m_{ik}(t_k) = Poiss(\lambda_i), \\ \lambda_i = N(1 - e^{\frac{-\mathbf{c}(t_k, \Theta_i) \times v}{N}}) \tag{9}$$

where N is the number of droplets, c is the concentration of 2-LTR circles in the blood, and v is the sample volume.

Using this simulated data we then use an MCMC technique to find posterior distributions for each parameter set $\Theta_i(A_i, \phi_i, R_i, \eta_{II}, \delta_i)$. The posterior distributions from [1] were the basis for the uninformative prior distributions $P(\Theta_i)$. The likelihood function for our MCMC calculation takes the form

$$\mathcal{L}(\Theta_i | m_{ik}) = f_{\mathcal{LN}}\big(m_{ik}(t_k), N(1 - e^{\frac{-\mathbf{c}(t_k, \Theta_i) \times v}{N}})\big) \tag{10}$$

where $f_{\mathcal{LN}}$ denotes the probability mass function of the Poisson distribution function [18].

Likewise for qPCR we assume measurement noise consistent with the assay which leads to measurements as

$$m_{ik} = lnN(ln(c(t_k, \Theta)), \ln(10)\sigma(n)) \tag{11}$$

Applying Bayes theorem we arrive at the equation

$$P(\Theta_i | m_{ik}) = \frac{\mathcal{L}(\Theta_i | mik) P(\Theta_i)}{\int_0^\infty P(m_{ik} | \Theta_i) P(\Theta) d\Theta} \tag{12}$$

However, $\int_0^\infty P(m_{ik} | \Theta_i) P(\Theta) d\Theta$ is a constant scaling factor of the posterior distribution [18]. For computational simplicity we simplify and arrive at the equation

$$P(\Theta_i | m_{ik}) \propto \mathcal{L}(\Theta_i | mik) P(\Theta_i) \tag{13}$$

which has the same form and conserves the KLD [12], [17].

F. Kullback-Leibler Divergence (KLD) Calculation

In order to quantify the fitness of each candidate schedule we calculate the Kullback-Leibler Divergence (KLD) between the prior multivariate distribution of the model parameters and the posterior distribution of the model parameters. Posterior distributions are calculated using a MCMC Metropolis-Hasting technique with Gibbs sampling [3]. The multivariate distribution is constructed from a set of five system parameters $\Theta(A, \phi, R, \eta_{II}, \delta)$. Parameters R, ϕ, η_{II}, and δ are exactly established from equation (2). Parameter A was derived as a an observable parameter which reduces the covariance between other parameters [1].

$$A \equiv \frac{k_{II} \mathbf{y_e} R}{\delta} \tag{14}$$

Calculation of the Kullback Leibler Divergence (KLD) is done using equation

$$KLD = \frac{1}{2}\Bigg(ln\Big(\frac{det\, \Sigma_2}{det\, \Sigma_1}\Big) - n - tr(\Sigma_2^{-1}\Sigma_1) \\ + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) \Bigg) \tag{15}$$

where (μ_1, Σ_1) and $(\mu 2, \Sigma_2)$ are the mean vector and covariances matrices of the prior and posterior multivariate distributions respectively and n is the number of dimensions in the distribution [12]. Because the natural log is used in the calculation of the KLD, the result is measured in natural units of information (nats). Equation 15 is applicable when all of the parameters are normally distributed. log(A),log(ϕ),and log(δ) are normally distributed. Parameters η_{II} and R are transformed using the normal distribution quantile function. The KLD between distributions is conserved through all transformations [12], [17].

G. Genetic Algorithm (GA)

Two recent integrase-intensification experiments have been performed: one by Hatano et al. at UCSF[3] and one by Lewin et al. at the University of Melbourne [24]. Hatano et al. took four samples at 0, 7, 14 and 56 days and Lewin et al. took samples at 0, 1, 3, 7, 14, 28, 56, and 84 days [3], [24]. With that in mind we intend to investigate whether or not their experiments can be improved upon, with respect to information gain, simply by altering the days at which the samples are taken. In order to locate a suboptimal sampling schedule we use a genetic algorithm (GA) to search the space of possible sampling schedules using Kullback-Liebler Diveregence as the fitness function [13], [14], [15].

The inherent binary nature of time series measurements, taking a sample on a given day or not in this case, lend themselves well to a genetic algorithm optimization method. For a fair comparison to the Hatano schedule we will use the GA to find a suboptimal four day sample schedule over a period of 87 days. This results in over 1e6 possible sample schedules. We will also use the GA to find a suboptimal eight day schedule over the same time period for comparison to the Lewin schedule which a search space of over 5.8e10

sample schedules. To perform and exhaustive calculation of all of the different possible schedules is not feasible, however the GA is able to quickly converge to a family of suboptimal solutions. To construct the GA, candidate

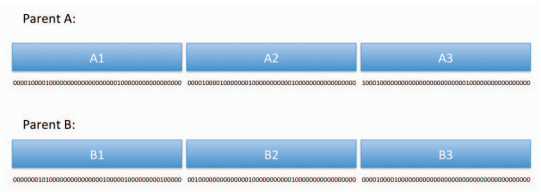

Fig. 1. Structure of Genes and Chromosomes in the Genetic Algorithm.

sample schedules are represented by a chromosome. Each chromosome consists of three genes and each gene is further broken down into 29 base pairs. Each base pair represents a potential sample day and takes on a binary value, 0 for days at which no sample and 1 for days at which a sample is taken. The three genes combine to form a chromosome with 87 base pairs or possible sample days as depicted in 1.

The algorithm is run with 20 child chromosomes per generation. The first generation is created by randomly selecting the appropriate number of base pairs (sample days) per chromosome: 4 for the Hatano GA and 8 for the Lewin Ga. The corresponding information content is then calculated for each chromosome by calculating its associated KLD. The chromosomes are then ranked in terms of the relative fitness by assigning chromosomes yielding higher KLD values a greater fitness level. The two chromosomes with the highest fitness are then used as the parent solutions to create the children for the next generation. Children are created through a process of genetic crossovers and mutations [13], [14], [15].

Genes are able to crossover to different locations or to the same location between parents. Point mutations occur after the crossovers to ensure that the chromosome has exactly four sample points. These mutations occur by bit inversion. For the 4 sample GA, if there are less than 4 sample points random 0s are inverted to 1s until 4 samples are achieved. Likewise, if there are more than 4 sample points 1s are inverted to 0s until 4 samples are achieved. If there are exactly 4 samples after the crossovers then two random inversions will occur to change a random 1 to a 0 and a random 0 to a 1. This is to ensure that the algorithm is able to escape local minima. The same is true in the 8 sample GA. After the first generation we employ an elite selection method, where the chromosome with the highest fitness throughout the evolution process is a parent from which to create children for all following generations. This ensures that solution quality is conserved throughout the evolutionary process.

III. RESULTS

A. Experiment Design Optimization for qPCR Assay

The first step in the optimization process is to run the Genetic Algorithm to locate a family of sub optimal sample schedules. Results of the GA runs for optimization using a qPCR Assay using 4 points and 8 points are shown in figure 2(A) and 2(B) respectively.

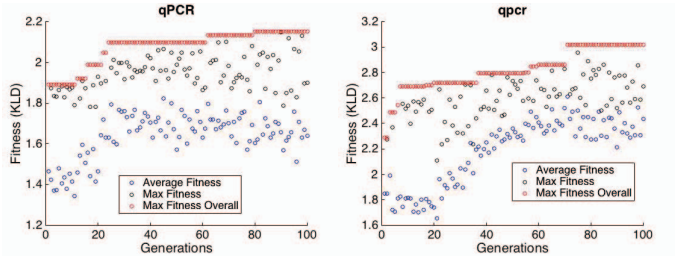

Fig. 2. (A) Genetic Algorithm Run for qPCR Assay with Four Sample Points (B) Genetic Algorithm Run for qPCR Assay with Eight Sample Points

Both runs of the GA were ran for 100 generations. The four sample optimization converged to a family of solution after around 40 generations while the the eight sample optimization converged to a family of solutions after around 60 generations. Another pronounced distinction between the two is that the eight sample GA run converged to a higher fitness level than the four sample. This is to be expected since more sample yield a greater gain in information from the experiment.

B. Experiment Design Optimization for ddPCR Assay

The plots in Fig. 3(A) and Fig. 3(B) show the results of the GA run for 4 sample optimization and 8 sample optimization respectively. Again we notice that the GA converges to a family of solutions that is slightly higher information for the 8 sample run.

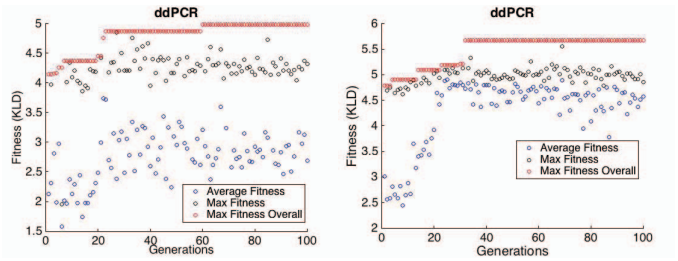

Fig. 3. (A) Genetic Algorithm Run for ddPCR Assay with Four Sample Points (B) Genetic Algorithm Run for ddPCR Assay with Eight Sample Points

C. Sample Point Distributions

To gain more insight into the differences between the GA optimization data we plot the distribution of the top 50 schedules, with respect to fitness, from each run.

From figure 4(A) we see that similar time points are chosen for both qPCR and ddPCR optimization. Particularly, points at the very beginning as well as sample points around day 26 are chosen. When the total number of samples is increased from 4 to 8 these two regions are chosen again as observed in figure 4(B). We also begin to see a pattern of samples being chosen around sample days 16 to 18. Samples are also consistently chosen later in the experiment, though the exact timing seems to be less important.

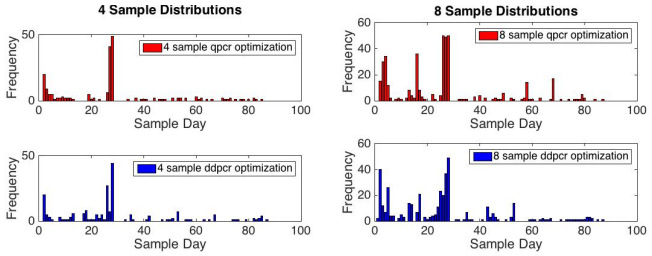

Fig. 4. (A) Distribution of the top 50 performing schedules from the 4 sample GA runs for both qPCR (red) and ddPCR (blue). (B) Distribution of the top 50 performing schedules from the 8 sample GA runs for both qPCR (red) and ddPCR (blue).

IV. Schedule Comparison

In order to test the validity of the optimization technique we compare a schedule from the family of solutions optimized for ddPCR to a schedule from a family of solutions optimized for qPCR and a previously used schedule from the experiment done by Hatano et al. The two plots in figure

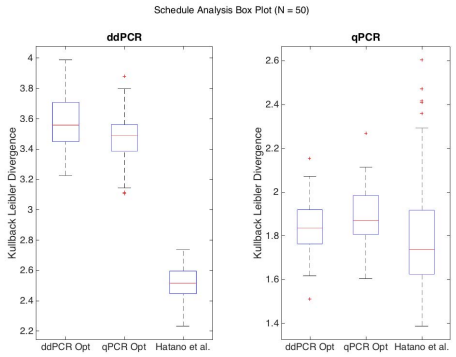

Fig. 5. Comparison of three 4-sample schedules: one optimized for ddPCR, one optimized for qPCR, and the 4-Sample Hatano et al. schedule

5 show the results of comparing the schedules. The plot on the left shows The three schedules analyzed with ddPCR noise. The schedule optimized for a ddPCR assay provides a slightly greater information gain than the schedule optimized for a qPCR assay and they both provide significantly data than the schedule used by Hatano et al. When analyzed with qPCR noise the schedule optimized for qPCR assay performs slightly better than the schedule optimized for a ddPCR assay; however, in this case neither of the two schedules are significantly better than the schedule used by Hatano et al. The plots in figure 7 display the results of the 8 point sample optimization. The plot on the left compares four sample schedules analyzed with ddPCR assay noise. The four schedules are a schedule chosen from the ddPCR optimized family of solutions, a schedule chosen form the qPCR optimized family of solutions, the schedule used in the experiment done by Lewin et al, and a uniform schedule that takes samples once per week.

The plot shows that both the ddPCR optimized and the qPCR optimized schedules perform very well. In addition, the information gained from the sample schedule used in

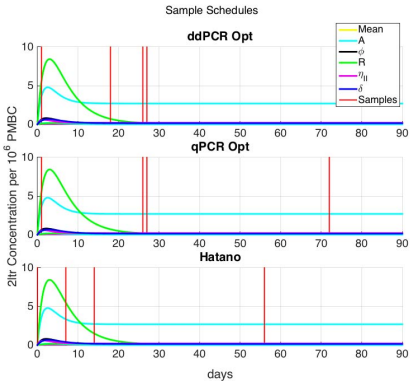

Fig. 6. Sample schedules from 4 sample optimization shown with the trajectories of the 11 sigma point patients

the Lewin is comparable to the optimized schedules. On the other hand the Uniform sampling schedule performs statistically significantly worse than the other three. An interesting finding is that the 4 sample schedules optimized for both ddPCR and qPCR from figure 5 yield a greater gain in information that the 8 sample uniform sample schedule. This demonstrates the importance of performing experiment design optimization methods such as those established in this paper. The right side of Figure 10 compares four

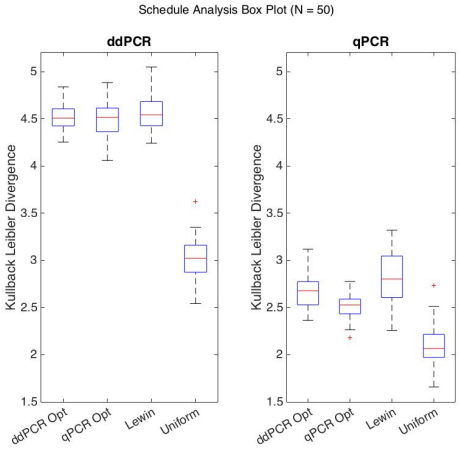

Fig. 7. Comparison of four 8-sample Schedules: one optimized for ddPCR, one for qPCR, an 8-sample schedule from the experiment done by Lewin et al. and a uniform schedule which samples once per week.

sample schedules analyzed with qPCR assay noise. The four schedules are the same as the plot on the left. The most obvious difference is the how much less information is gained when the schedules are analyzed with qPCR assay noise. The poorly performing Uniform schedule from the ddPCR assay out performs all of the schedules analyzed with qPCR noise. Of the schedules analyzed with qPCR noise the optimized schedules and the Lewin schedules once again perform better than the Uniform schedule. The schedule used in the experiments done by Lewin et al. actually perform better than the two optimized schedules that were selected. These results would imply that when designing an

experiment using a qPCR assay selection of sample schedule has less of an effect on the overall amount of information gained in the experiment. This is due to the high level of noise associated with the qPCR assay. Figure 8 shows the

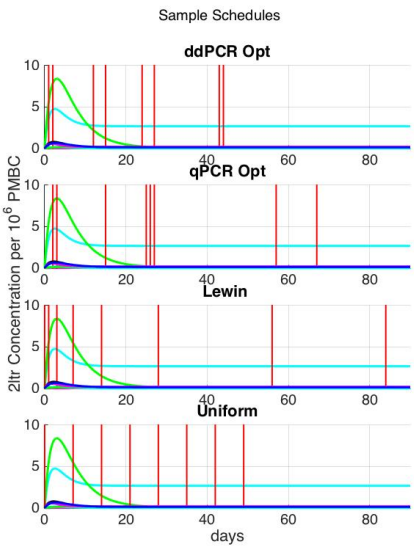

Fig. 8. Sample schedules from 8-sample optimization shown with the trajectories of the 11 sigma-point patients

location of the sample points (vertical red line) plotted over the the trajectories of the 11 sigma point patients.

V. CONCLUSION

In this paper we have demonstrated the advantages of using KLD in the design of clinical experiments. Our method has employed the use of a genetic algorithm to find sub-optimal sample schedules based on the measurement assay to be used. Analysis of the resulting schedules from the GA when compared to other schedules show that there is a need to both optimized sample schedule and take into consideration what type of measurement assay is being used when designing an experiment. Because ddPCR provides a much more accurate measurement our analysis indicates that using an optimization method is of utmost importance when using this assay. When designing an experiment using ddPCR noise, it would benefit the researcher to optimize the sampling schedule to obtain as much information as possible. Our results fit well in the literature on pratical identifiability for biomedical systems [25].

ACKNOWLEDGMENTS

We would like to thank Sharon Lewin and Hiroyu Hatano for sharing their knowledge of HIV clinical trial design.

REFERENCES

[1] Luo, Rutao, et al. "Modelling HIV-1 2-LTR dynamics following raltegravir intensification". Journal of the Royal Society Interface 10.84 (2013): 20130186.

[2] Luo, Rutao, et al. "HIV model parameter estimates from interruption trial data including drug efficacy and reservoir dynamics." PLoS One 7.7 (2012): e40198.

[3] Hatano, Hiroyu, et al. "Increase in 2long terminal repeat circles and decrease in D-dimer after raltegravir intensification in patients with treated HIV infection: a randomized, placebo-controlled trial". Journal of Infectious Diseases 208.9 (2013): 1436-1442.

[4] Hatano, Hiroyu, et al. "Pre-HAART HIV burden approximates post-HAART viral levels following interruption of therapy in patients with sustained viral suppression." Aids 14.10 (2000): 1357-1363.

[5] Chun, Tae-Wook, et al. "Relationship between pre-existing viral reservoirs and the re-emergence of plasma viremia after discontinuation of highly active anti-retroviral therapy." Nature medicine 6.7 (2000): 757-761.

[6] Chun, Tae-Wook, et al. "AIDS: re-emergence of HIV after stopping therapy." Nature 401.6756 (1999): 874-875.

[7] Davey, Richard T., et al. "HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression." Proceedings of the National Academy of Sciences 96.26 (1999): 15109-15114.

[8] Wong, Joseph K., et al. "Recovery of replication-competent HIV despite prolonged suppression of plasma viremia." Science 278.5341 (1997): 1291-1295.

[9] Persaud, Deborah, et al. "Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads." Journal of virology 78.2 (2004): 968-979.

[10] Maldarelli, Frank, et al. "ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia." PLoS Pathog 3.4 (2007): e46.

[11] Dornadula, Geethanjali, et al. "Residual HIV-1 RNA in blood plasma of patients taking suppressive highly active antiretroviral therapy." Jama 282.17 (1999): 1627-1632.

[12] Duchi, John. "Derivations for linear algebra and optimization." Berkeley, California (2007).

[13] Tsujimura, Yasuhiro, and Mitsuo Gen. "Entropy-based genetic algorithm for solving TSP." Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES'98. 1998 Second International Conference on. Vol. 2. IEEE, 1998.

[14] Bäck, Thomas, and Hans-Paul Schwefel. "An overview of evolutionary algorithms for parameter optimization." Evolutionary computation 1.1 (1993): 1-23.

[15] Sivanandam, S. N., and S. N. Deepa. Introduction to genetic algorithms. Springer Science and Business Media, 2007.

[16] Julier, Simon J., and Jeffrey K. Uhlmann. "Unscented filtering and nonlinear estimation." Proceedings of the IEEE 92.3 (2004): 401-422.

[17] Kullback, Solomon, and Richard A. Leibler. "On information and sufficiency." The annals of mathematical statistics (1951): 79-86.

[18] Luo, Rutao, Michael J. Piovoso, and Ryan Zurakowski. "Modeling uncertainty in single-copy assays for HIV." Journal of clinical microbiology 50.10 (2012): 3381-3382.

[19] Buzon, Maria J., et al. "HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects." Nature medicine 16.4 (2010): 460-465.

[20] Strain, Matthew C., et al. "Highly precise measurement of HIV DNA by droplet digital PCR." PloS one 8.4 (2013): e55943.

[21] Cannon, LaMont, et al. "Prospective HIV clinical trial comparison by expected Kullback-Leibler Divergence." American Control Conference (ACC), 2016. IEEE, 2016.

[22] Abraham, George, et al. "Order preservation of expected information content using Unscented Transform approximation of multivariate prior distributions in HIV 2-LTR experiment design." Decision and Control (CDC), 2016 IEEE 55th Conference on. IEEE, 2016.

[23] Vandergeeten, Claire, et al. "Cross-clade ultrasensitive PCR-based assays to measure HIV persistence in large-cohort studies." Journal of virology 88.21 (2014): 12385-12396.

[24] Lewin, Sharon; University of Melbourne; The Alfred;ViiV Healthcare. Dolutegravir Impact on Residual Replication: Dolutegravir Intensification Study. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000-[cited 2017 March 20]. Available From: https://clinicaltrials.gov/show/NCT02500446.

[25] Nguyen, Van Kinh et al "Analysis of Practical Identifiability of a Viral Infection Model." PLoS ONE 11.12 (2016): e0167568