



GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH



Quantifying the HIV Reservoir with Dilution Assays and Deep Viral Sequencing

Brian Richardson

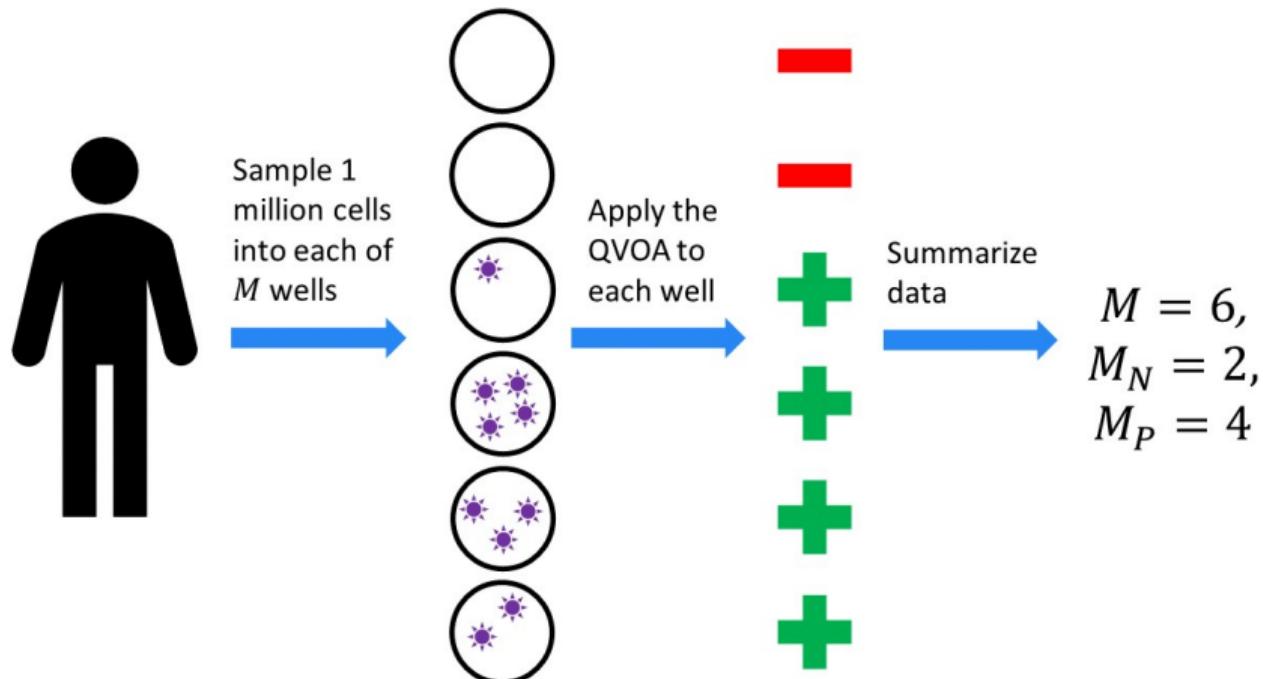
Motivation: Why Quantify the HIV Reservoir?

- For people living with HIV, antiretroviral therapies can help them achieve **viral suppression**
- Despite viral suppression, a **reservoir of latently infected cells**, which are undetectable by the immune system, remain
- The HIV reservoir is often measured in **infectious units per million**, or **IUPM**
- A barrier in HIV cure research is the need to **reliably quantify the IUPM** of the HIV reservoir

Quantitative Viral Outgrowth Assay

- One standard method for quantifying the HIV reservoir is a **serial limiting dilution (SLD) assay** called the **Quantitative Viral Outgrowth Assay**, or **QVOA**
- Cells are sampled from one individual into multiple wells
- The QVOA tests each well for the presence of **at least one HIV-infected cell**

Quantitative Viral Outgrowth Assay: Example



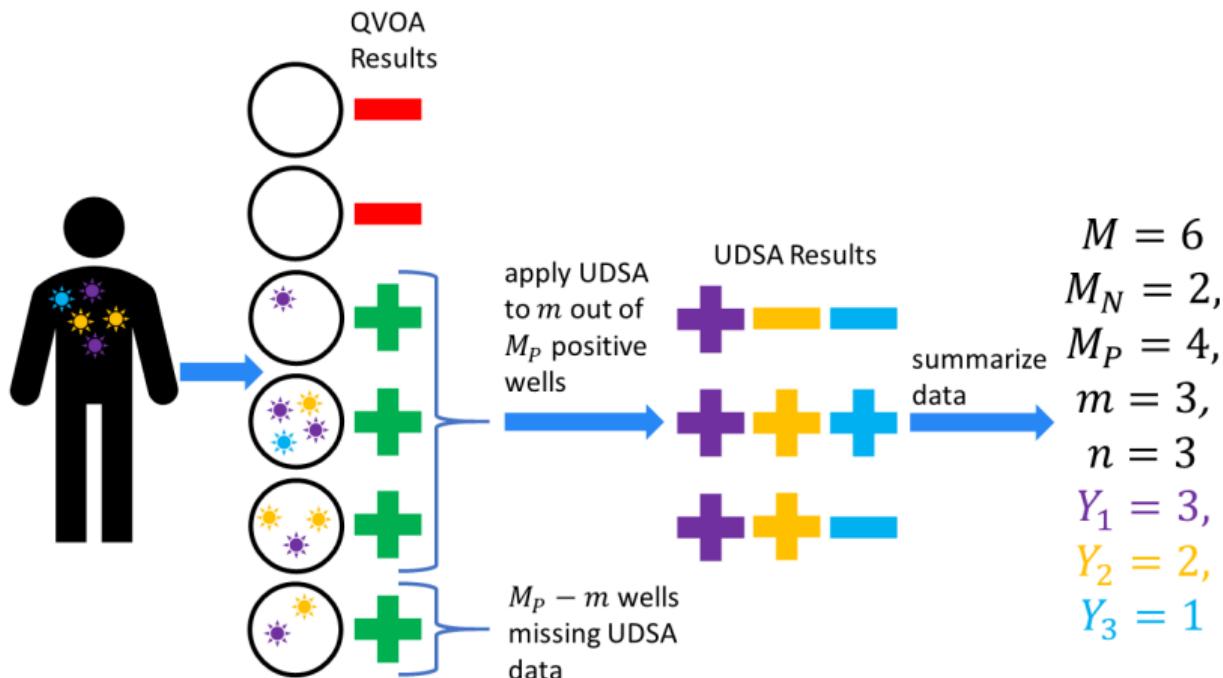
Quantitative Viral Outgrowth Assay: Inference

- latent cell counts $X_1, \dots, X_M \sim \text{Poisson}(\Lambda)$
- observed indicators $W_j = 1(X_j > 0)$
- estimand: IUPM $\Lambda = E(X_j)$
- MLE for $E(W_j) = 1 - \exp(-\Lambda)$ is sample mean $\frac{M_P}{M}$
- MLE for Λ is then $\hat{\Lambda} = -\log(1 - \frac{M_P}{M})$

Ultra Deep Sequencing Assay

- The Ultra Deep Sequencing Assay, or UDSA, is a more sophisticated assay used in addition to the QVOA
 - typically applied to wells that the QVOA has identified as HIV-positive
 - tests for the presence of **distinct viral lineages (DVLs)**

Ultra Deep Sequencing Assay: Example



Inference about the IUPM

- Let λ_i be the **DVL-specific IUPM** for DVL i and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$
- Let $\Lambda = \sum_{i=1}^n \lambda_i$ be the **IUPM** (over all DVLs of HIV)
- Our goal is to estimate Λ with $\widehat{\Lambda} = \sum_{i=1}^n \widehat{\lambda}_i$
- With no missing data, MLE for λ_i is $\widehat{\lambda}_i = -\log(1 - \frac{Y_i}{M})$
- Not so easy with missing data...

Inference about the IUPM

The **observed-data likelihood** is $L(\boldsymbol{\lambda} | M_N, \mathbf{Y}) =$

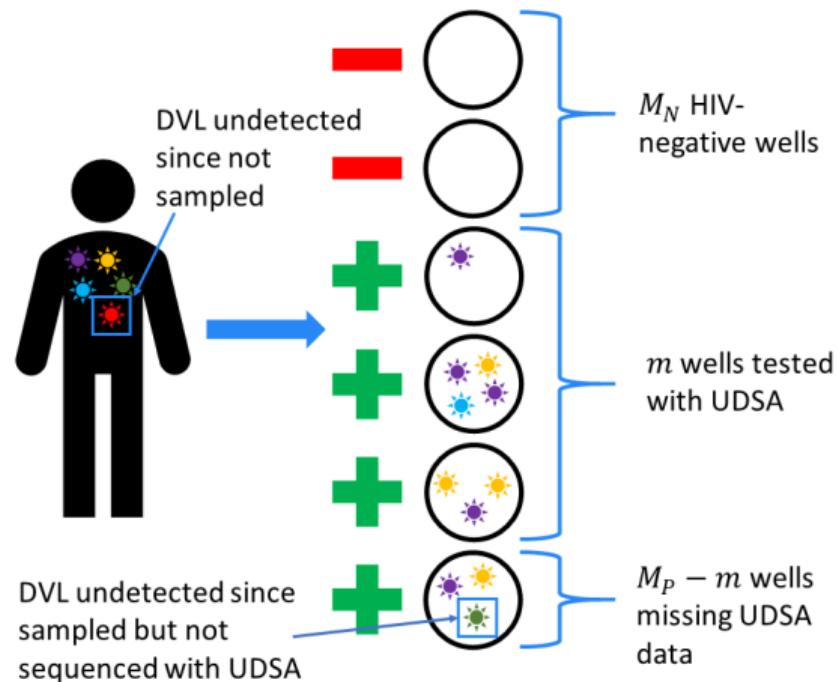
$$\underbrace{\left[\prod_{i=1}^n \{1 - \exp(-\lambda_i)\}^{Y_i} \exp(-\lambda_i)^{M_N+m-Y_i} \right]}_{\text{complete data}} \underbrace{\{1 - \exp(-\Lambda)\}^{M-(M_N+m)}}_{\text{missing data}}$$

- Maximized numerically to get the MLE $\hat{\boldsymbol{\lambda}}$
- The **MLE for the IUPM** is $\hat{\Lambda} = \sum_{i=1}^n \hat{\lambda}_i$
- $\hat{\boldsymbol{\lambda}}$ (and thus $\hat{\Lambda}$) is consistent and asymptotically normal

Bias Correction for Small Samples

- The MLE $\hat{\lambda}$ is **upwardly biased** in small samples (i.e., small M)
- A **bias-corrected MLE** $\hat{\lambda}^*$ is adapted from Hashemi & Schneider
PLoS ONE 2021
- $\hat{\lambda}^* = \hat{\lambda} - B(\hat{\lambda})$, where $B(\hat{\lambda})$ is an estimate of the bias of $\hat{\lambda}$
- This bias correction reduces the order of the bias from $\mathcal{O}(M^{-1})$ to $\mathcal{O}(M^{-2})$

Undetected Viral Lineages: Example



- Existing DVLs:
 $n' = 5$
- Detected DVLs:
 $n = 3$
- True IUPM: $\Lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5$
- Estimated IUPM:
 $\widehat{\Lambda} = \widehat{\lambda}_1 + \widehat{\lambda}_2 + \widehat{\lambda}_3$

Undetected Viral Lineages: Problem

- Recall: $\hat{\Lambda} = \hat{\lambda}_1 + \cdots + \hat{\lambda}_n$
- n is the number of **observed DVLs**. There may exist $n' > n$ DVLs in the source population, leaving $n' - n$ DVLs **undetected**
- Then $\Lambda = \lambda_1 + \cdots + \lambda_n + \lambda_{n+1} + \cdots + \lambda_{n'}$, which is greater than $\lambda_1 + \cdots + \lambda_n$
- Does this mean $\hat{\Lambda}$ is a poor estimator of Λ ?

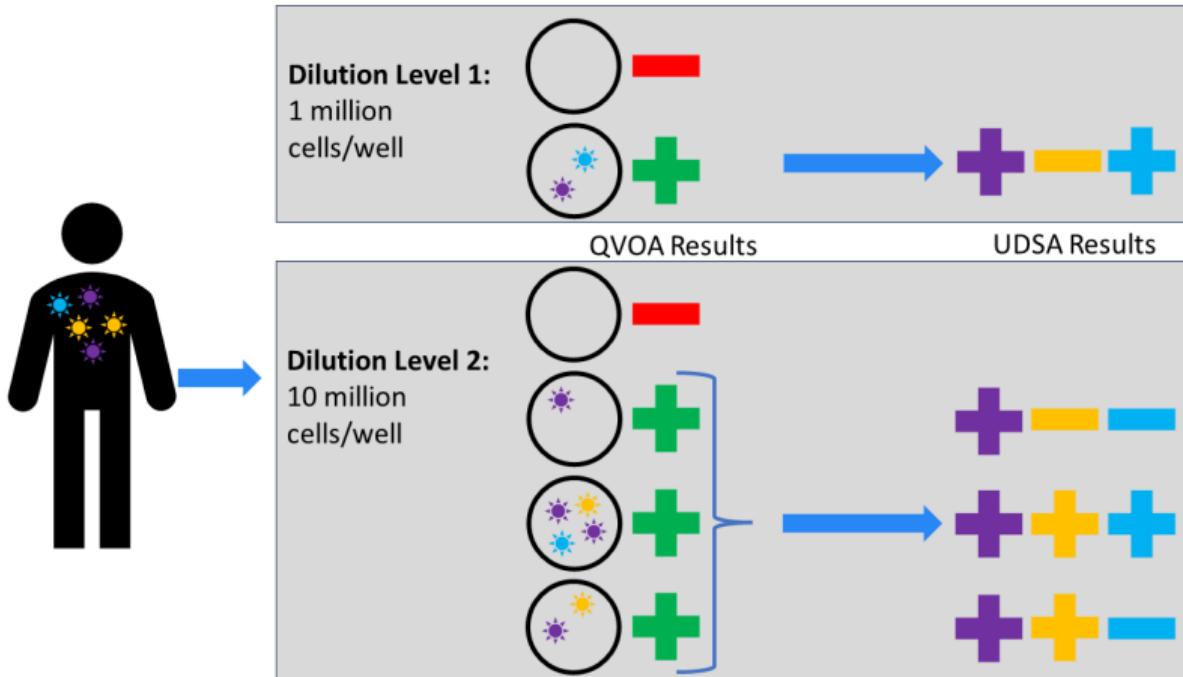
Undetected Viral Lineages: Solution

- Consider an **augmented likelihood** that accounts for possibly undetected DVLs

$$L'(\lambda'_1, \dots, \lambda'_n, \lambda'_{n+1}, \dots, \lambda'_{n'} | M_N, Y_1, \dots, Y_n, Y_{n+1} = 0, \dots, Y_{n'} = 0)$$

- The maximizer $\hat{\lambda}'$ of the augmented likelihood L' must satisfy $\hat{\lambda}'_{n+1} = 0, \dots, \hat{\lambda}'_{n'} = 0$
- Thus, $\hat{\Lambda} = \sum_{i=1}^n \hat{\lambda}_i = \sum_{i=1}^{n'} \hat{\lambda}'_i$
- The augmented likelihood L' and the original likelihood L lead to the same MLE $\hat{\Lambda}$

Multiple Dilution Levels: Example



Extension to Multiple Dilution Levels

- Often, the QVOA and UDSA are applied to wells of **multiple dilution levels**
- Previous estimation methods can handle QVOA data from multiple dilutions or QVOA and UDSA data from one dilution level
- We proposed a more general estimator that can handle both
 - The bias-corrected MLE and the undetected DVL result extends to the multiple dilutions setting

Simulation Results

IUPM of $\Lambda = 1$, single dilution level, and 75% of HIV-positive wells tested with UDSA

| n' | M | MLE | | | Bias-Corrected MLE | | |
|------|-----|------|------|------|--------------------|------|------|
| | | Bias | ESE | CP | Bias | ESE | CP |
| 12 | 12 | 0.04 | 0.33 | 0.94 | -0.02 | 0.31 | 0.96 |
| 12 | 24 | 0.03 | 0.22 | 0.95 | -0.01 | 0.21 | 0.96 |
| 18 | 12 | 0.05 | 0.32 | 0.94 | -0.01 | 0.31 | 0.96 |
| 18 | 24 | 0.03 | 0.23 | 0.95 | 0.00 | 0.22 | 0.96 |

ESE: empirical standard error of $\hat{\Lambda}$,

CP: empirical coverage probability of 95% confidence interval

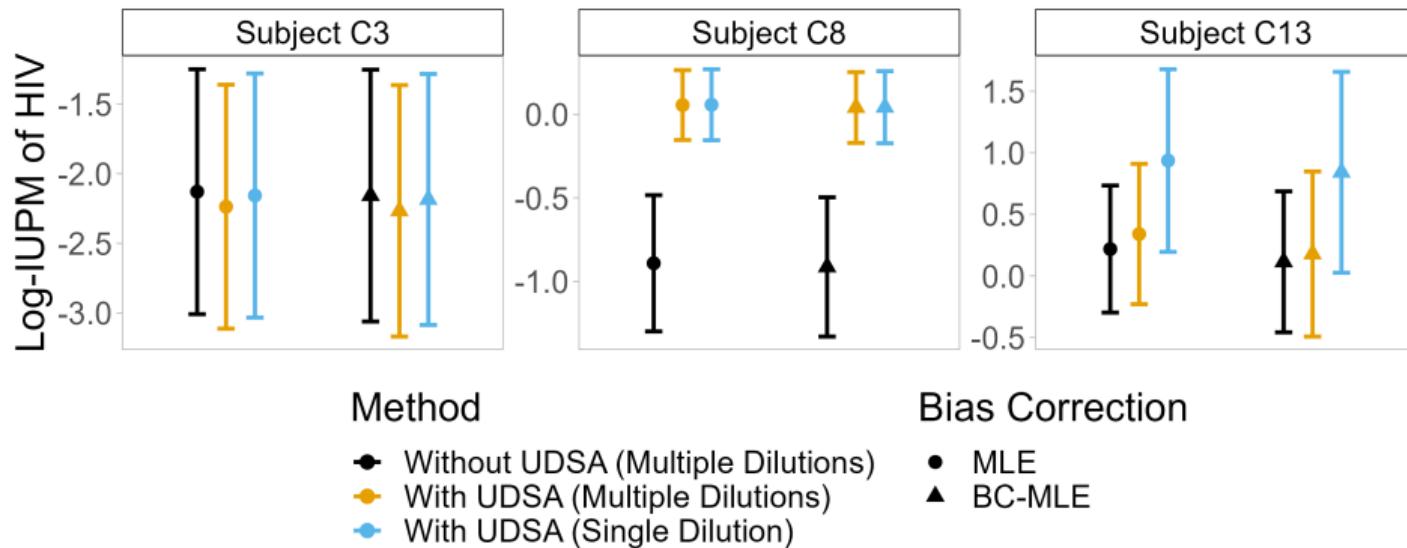
Application to HIV Data

QVOA and UDSA data on individuals living with HIV and taking antiretroviral treatment were obtained from the UNC HIV Cure Center

- For each individual, cells were sampled into various wells at **3–4 distinct dilution levels** and tested with the QVOA
- A subset of wells from **one dilution level was deep-sequenced**

| Subject ID | DLVs (n) | Wells (M) | Positive Wells (M_P) | UDSA Wells (m) |
|------------|--------------|---------------|--------------------------|--------------------|
| C3 | 4 | 18, 6, 6 | 5, 0, 0 | 5, 0, 0 |
| C8 | 26 | 36, 6, 6 | 22, 2, 1 | 22, 0, 0 |
| C13 | 7 | 18, 6, 6, 6 | 16, 4, 3, 0 | 0, 4, 0, 0 |

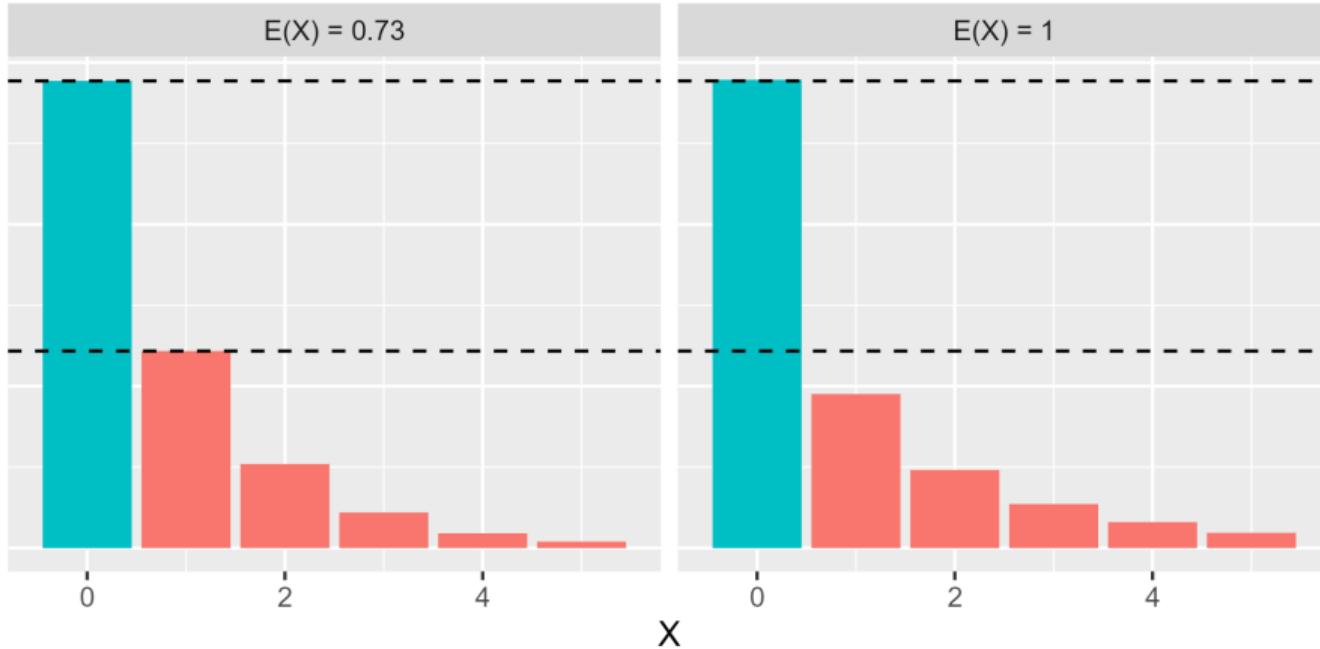
Application to HIV Data



Overdispersed Cell Counts: Problem

- Currently assuming latent cell counts are $X_i \sim \text{Poisson}(\lambda_i)$
- Allows inference on $\Lambda = E(X_i)$ based on observed $W_i = 1(X_i > 0)$
- Poisson assumption based on the approximation to $\text{Binomial}(n, p)$ with large n and small p .
- May be concern that Poisson assumptions doesn't hold
- **Overdispersion:** $\text{Var}(X_i) > E(X_i)$

Overdispersed Cell Counts: Example



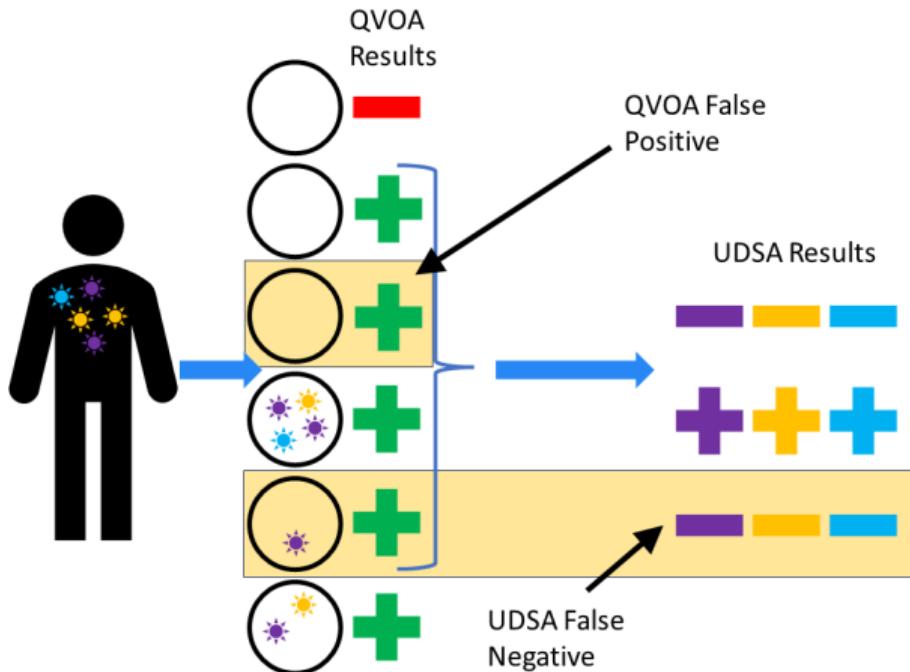
Overdispersed Cell Counts: Solution

- Model X_i as negative binomial with mean parameter λ_i and dispersion parameter $\gamma \in [0, \infty)$
- $P(X_i > 0) = 1 - (\gamma\lambda_i + 1)^{-1/\gamma}$
- When $\gamma = 0$, reduces to Poisson case
- With multiple dilution level data, λ, γ are identifiable
- Can estimate λ, γ
- Likelihood ratio test of overdispersion $H_0 : \gamma = 0$
 - null parameter on boundary of parameter space
 - asymptotic null distribution is $0.5\chi_0^2 + 0.5\chi_1^2$

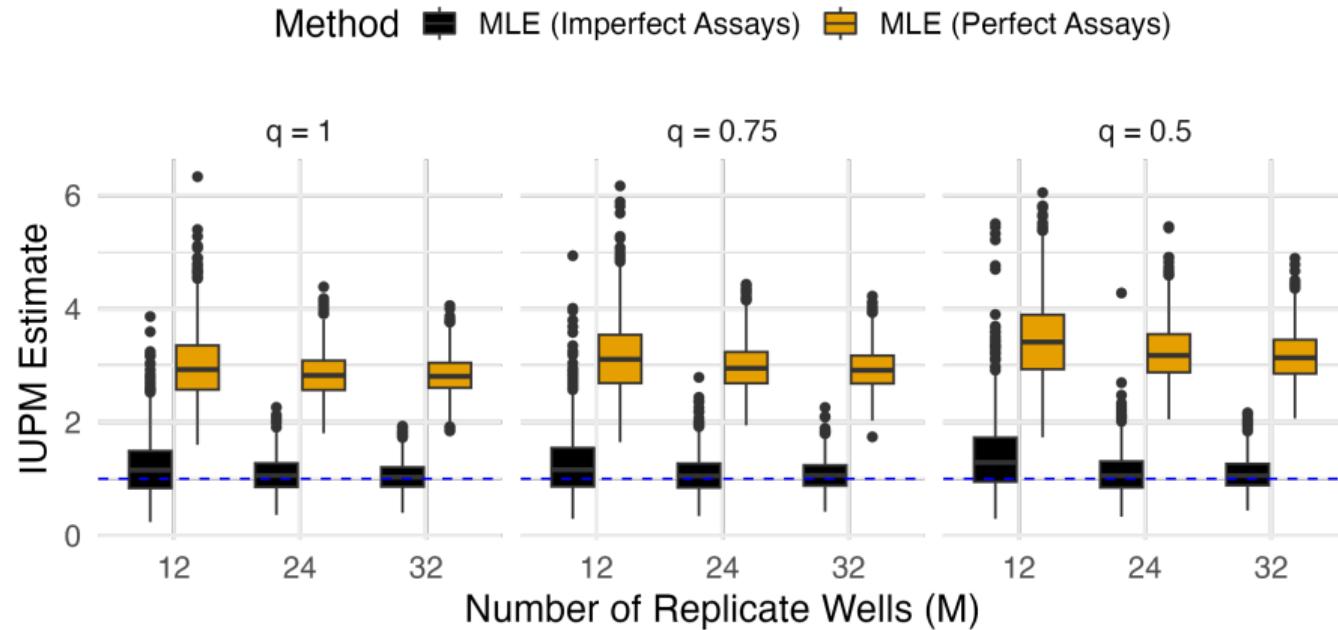
Imperfect Assays: Problem

- So far, we've assumed that the two assays (QVOA and UDSA) have perfect sensitivity and specificity
- **Sensitivity:** Probability of a positive result given the well is truly infected
- **Specificity:** Probability of a negative result given well is truly not infected
- Need to account for false positives and negatives in estimation of Λ

Imperfect Assays: Example

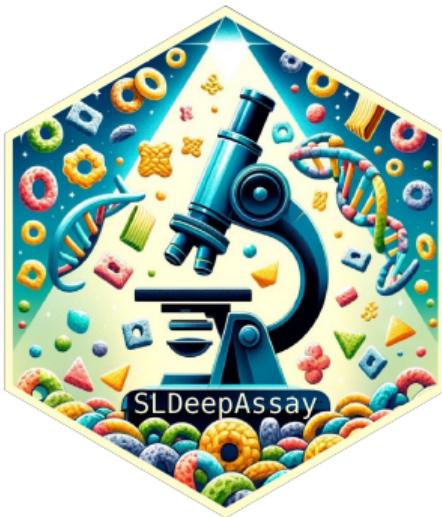


Imperfect Assays: Simulation Results



(single dilution level, 90% sensitivity and specificity of both assays)

Want to Learn More?



R package: <https://github.com/sarahlotspeich/SLDeepAssay>
Paper: <https://pubmed.ncbi.nlm.nih.gov/38364812/>

Acknowledgements

- Co-authors: Sarah C. Lotspeich, Pedro L. Baldoni, Kimberly P. Enders and Michael G. Hudgens
- Our collaborators at the UNC Center for AIDS Research (CFAR)
- This research was supported by the following NIH grants
 - T32ES007018
 - P30AI50410
 - R37AI029168

Thank you! Any questions?

Brian Richardson

✉: brichson@ad.unc.edu