Analysis of left-censored multiplex immunoassay data: A unified approach

Elizabeth G. Hill
Medical University of South Carolina

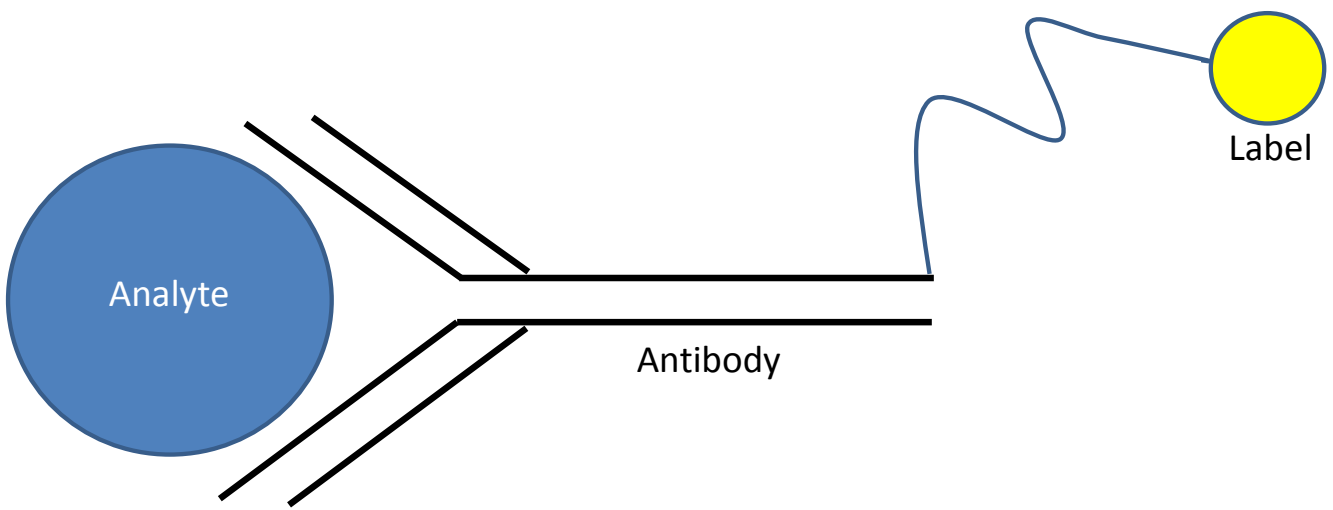
Elizabeth H. Slate
Florida State University

FSU Department of Statistics Seminar Series
20 September 2013

Talk Outline

An immunoassay is ...

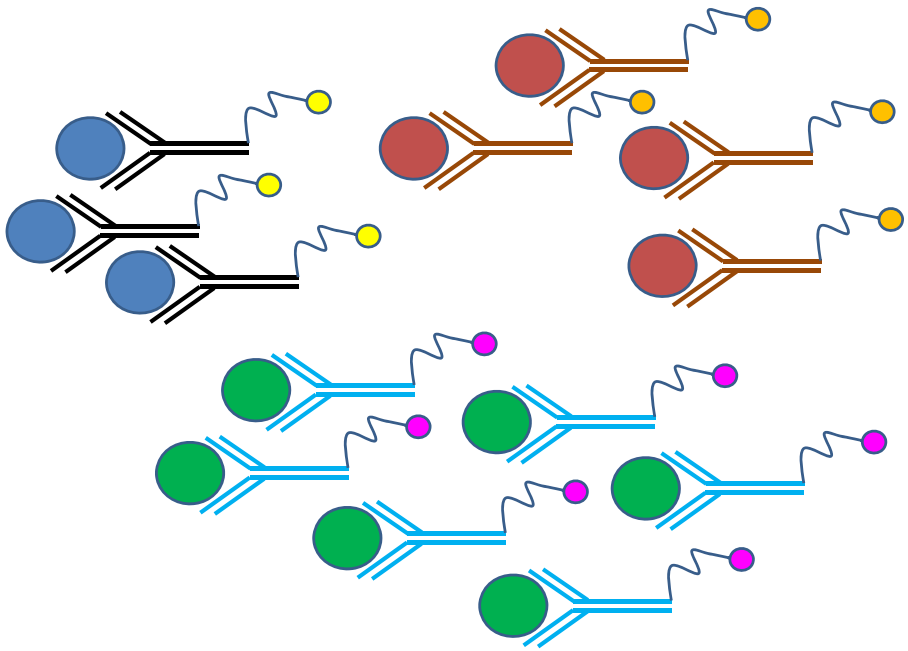
- A chemical test used to measure the concentration of a molecule in solution
- Target molecule - often a protein - called an *analyte*
- Analyte is “captured” by an analyte-specific antibody
- Measureable “label” (e.g. intensity of fluorescent signal) facilitates quantitation



Conventional uses

- Detect presence (absence) of a protein
  - Pregnancy test
  - Steroid use
  - Goal: *"Is analyte X detectable in this subject?"*
- Measure the concentration of protein
  - Blood panel
  - Viral load (HIV+ patients)
  - Goal: *"What is the concentration of analyte X for this subject?"*
- Emphasis is on analyte measurement for an individual

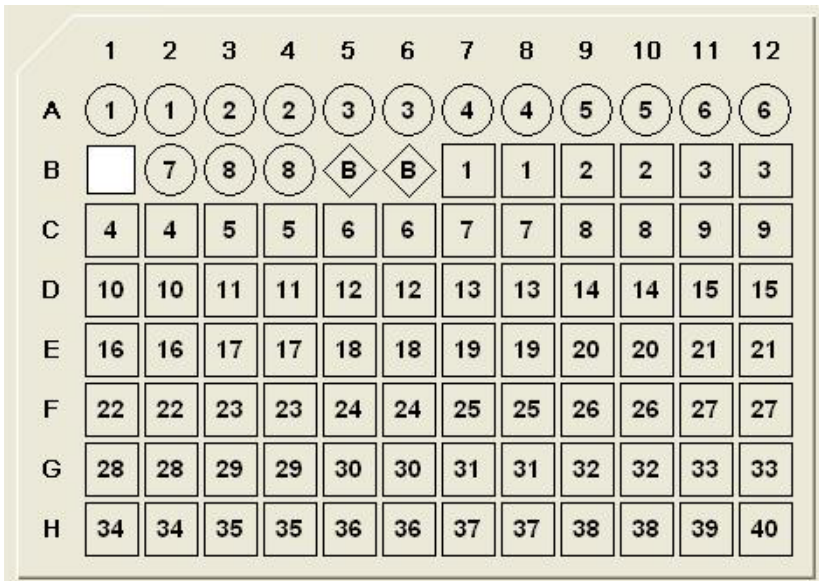
Multiplex immunoassay



- Simultaneous quantitation of a panel of analytes
- Commonly used as biomarker discovery tool
- Emphasis on group-level comparisons
- Goal
  - *"Is analyte X associated with disease status?"*
  - *"Does analyte X predict disease status?"*

Multiplex immunoassay

- All analytes accommodated on a single 96-well plate
- k-plex, $k = 5 - 30$ (typically)
- Multiple plates used to facilitate large sample sizes
- Each well detects all k analytes



Standards – serially diluted known analyte concentrations

Blanks – no analyte concentration

Patient samples – unknown analyte concentration

Workflow

For every analyte on each plate

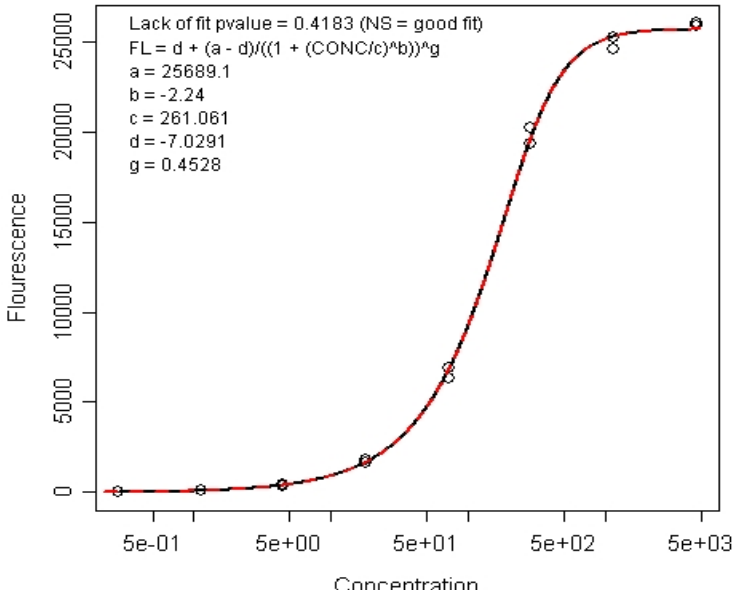
1. Place $8 - 10$ duplicate standard concentrations
2. Place duplicate blanks
3. Use remaining wells for duplicate patient samples (serum, urine, saliva)
4. Measure signal intensity (*response*)
5. Calculate average response for blanks
6. "Correct" standard and patient responses by subtracting the average blank response
7. Fit a standard curve to paired known standard concentrations and background corrected (BgC) responses
8. Average the patient BgC responses
9. Back-fit average BgC patient response values to obtain estimates of analyte concentrations
10. Conduct group-level inference

Standard curve

5-Parameter Logistic Model

$$f(x_{ijk}|\beta_{jk}) = d_{jk} + \frac{a_{jk} - d_{jk}}{\left[1 + \left(\frac{x_{ikj}}{c_{jk}}\right)^{b_{jk}}\right]^{g_{jk}}}$$

where x_{ijk} is the ith known concentration for analyte j plate k and $\beta_{jk} = (a_{jk}, b_{jk}, c_{jk}, d_{jk}, g_{jk})'$.

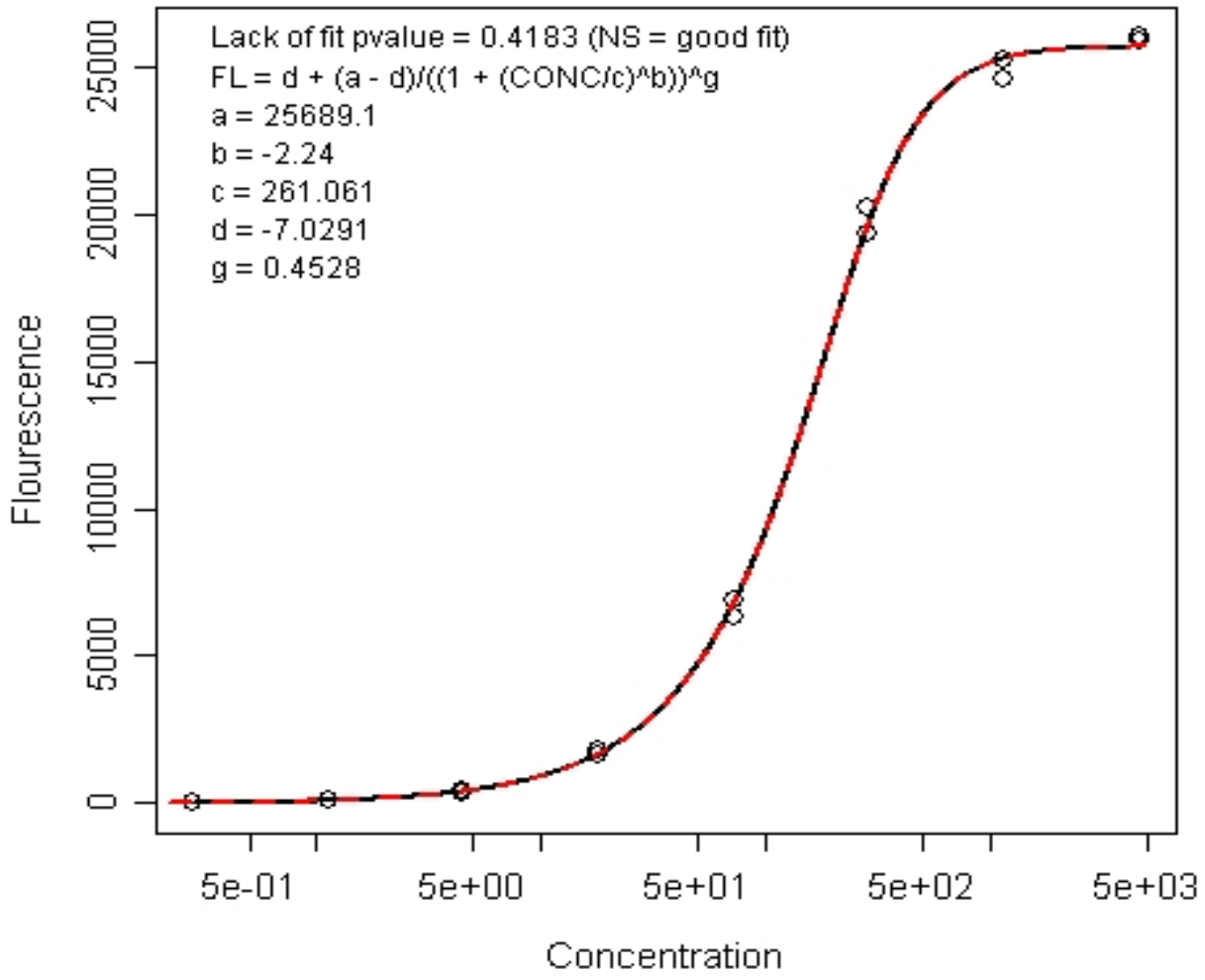


a_{jk} - upper asymptote
d_{jk} - lower asymptote
b_{jk} - transition rate between asymptotes
c_{jk} - concentration at inflection point
g_{jk} - asymmetry parameter

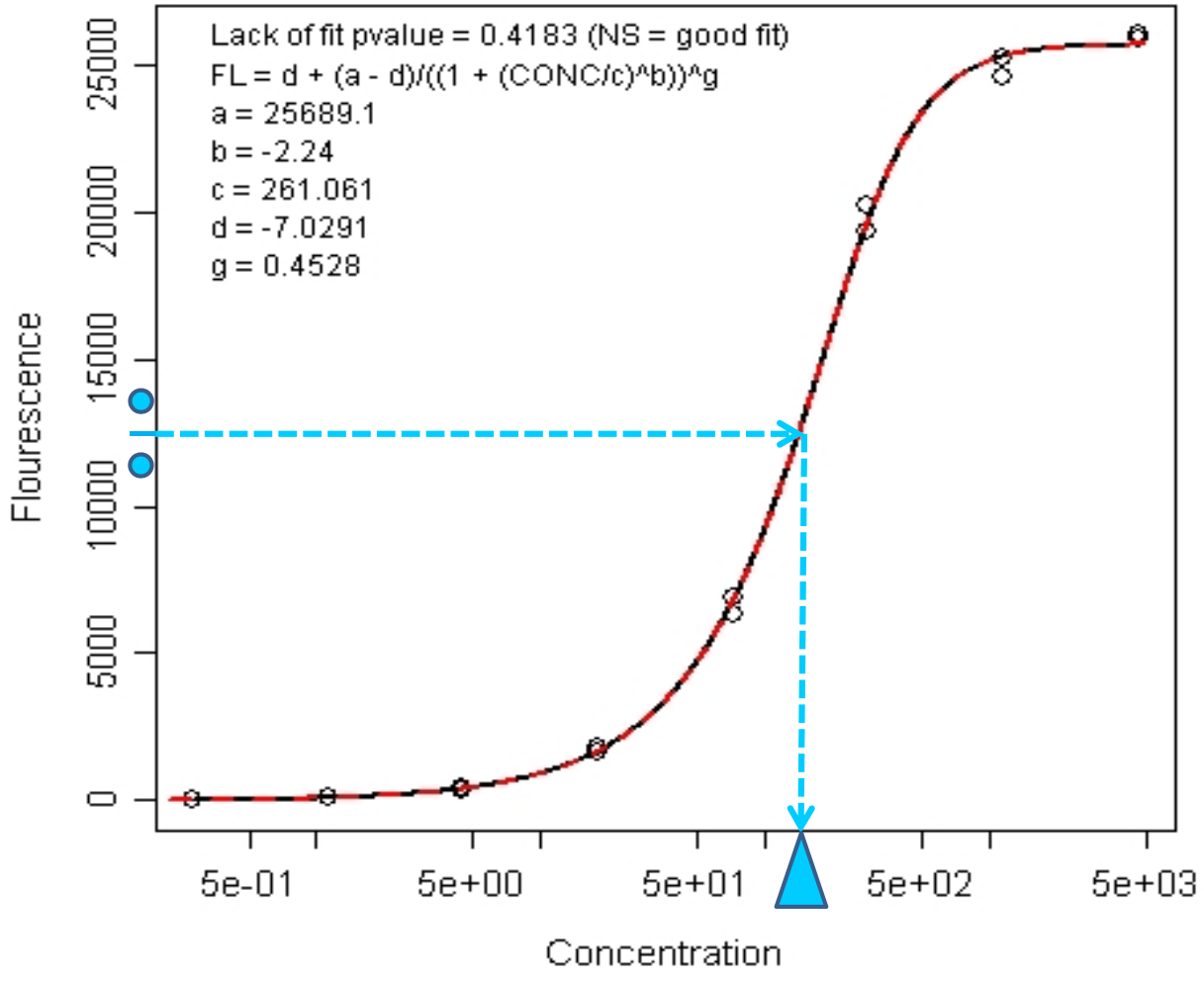
Fitting the 5-PL standard curve

- $y^*_{ijk\ell}$ = observed BgC response corresponding to ℓth replicate of standard concentration x_{ijk}
- $y^*_{ijk\ell} \sim \text{Normal}(f(x_{ijk}|\beta_{jk}), \text{Var}(y^*_{ijk\ell}))$
- Variance systematically related to mean
  - Noise in signal detectors is proportional to response magnitude
  - Kinetics associated with antibody binding
  - Power of the mean variance function
  - $\text{Var}(y^*_{ijk\ell}) = \sigma^2_{jk} f(x_{ijk}|\beta_{jk})^{2\theta}$
- Curve fitting achieved using generalized least squares (GLS)
- GLS = weighted least squares with weights estimated by $\text{Var}(y^*_{ijk\ell})^{-1}$

Standard curve



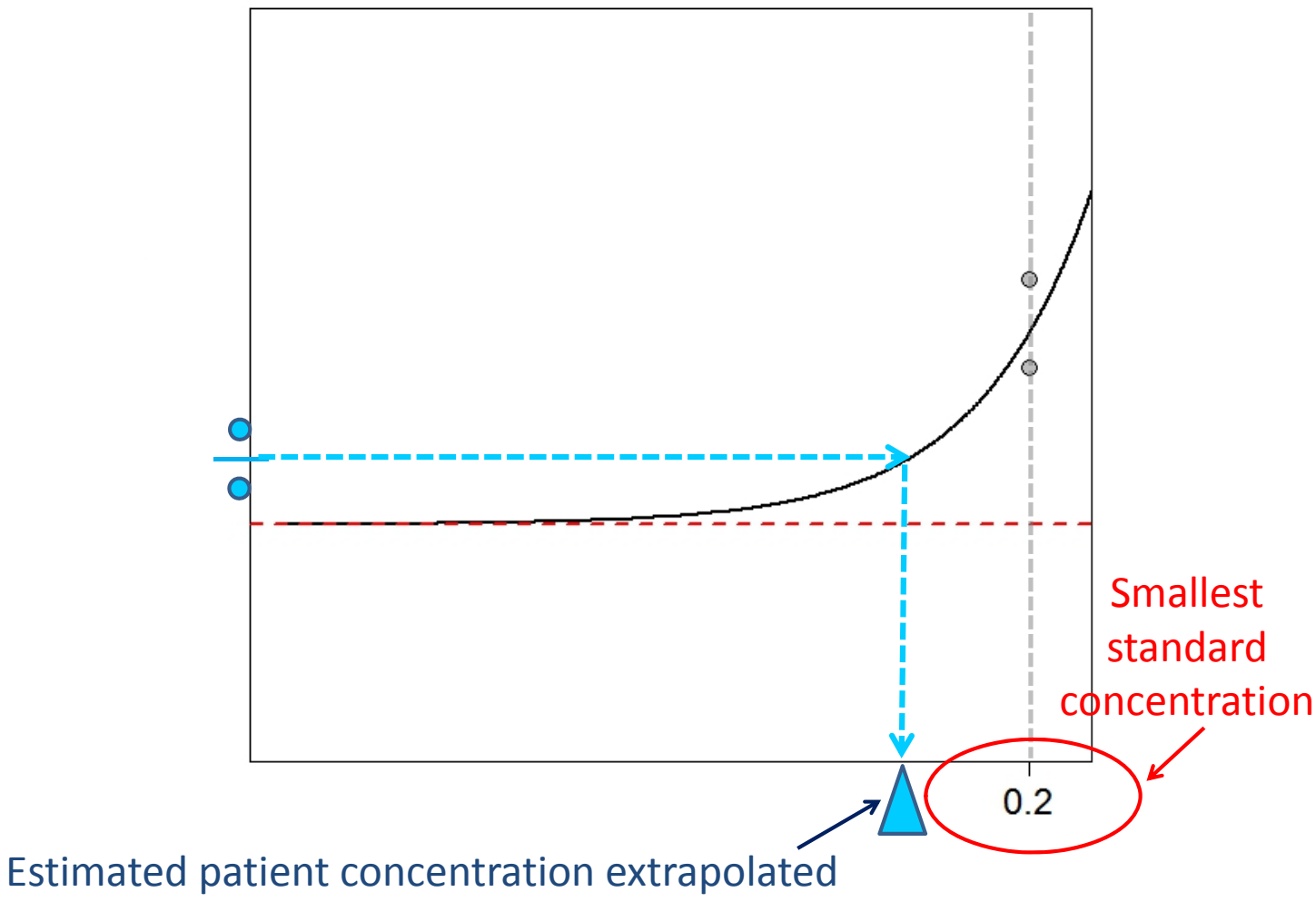
Patient analyte concentration estimation



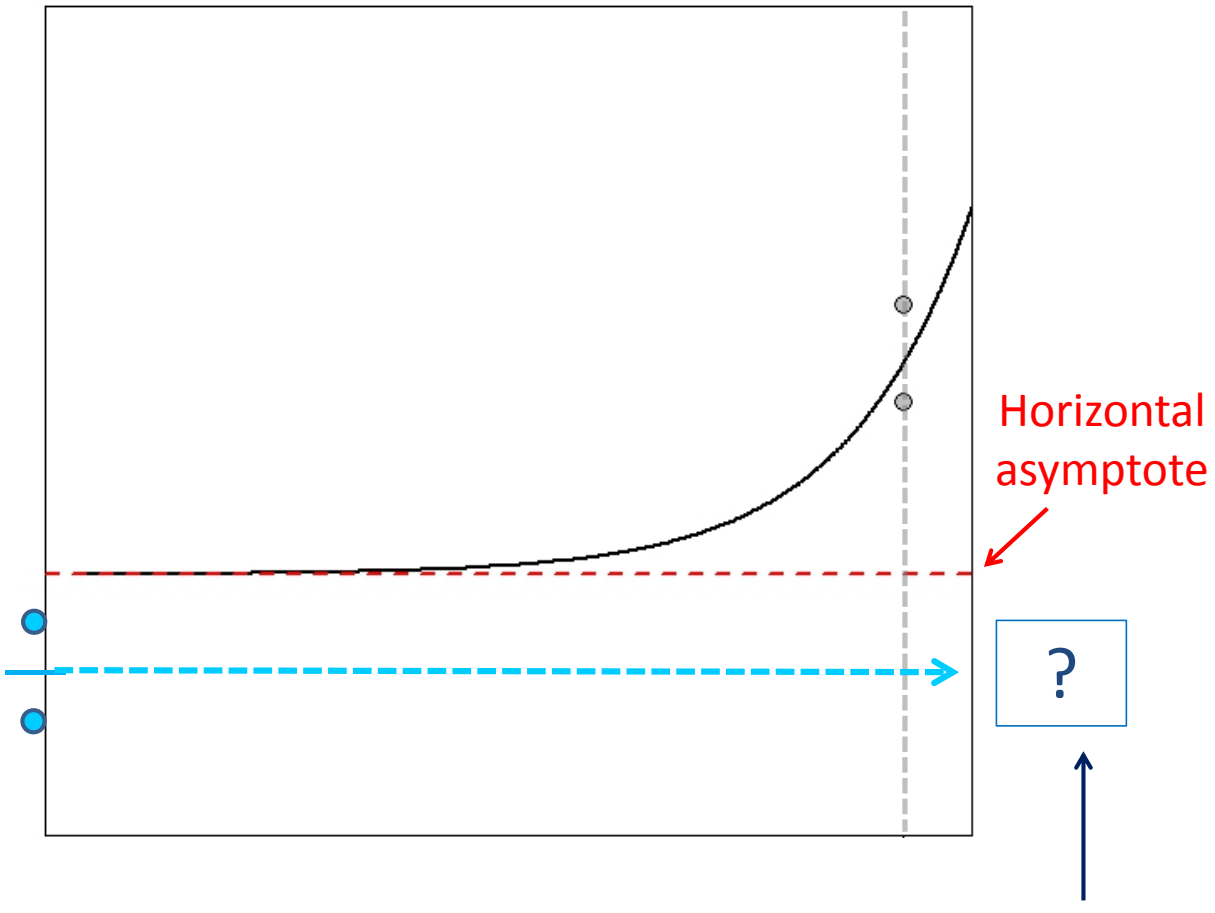
Sounds straightforward until you look at the data

Well	IL-1b	IL-2	IL-4	IL-5
B7,B8	*0.28	23.41	0.42	1.74
B9,B10	*0.61	51.61	1.07	3.94
B11,B12	OOR <	5.58	OOR <	OOR <
C1,C2	1.04	58.38	1.18	5.59
C3,C4	*0.79	63.17	1.66	5.69
C5,C6	*0.27	25.23	0.38	2.32
C7,C8	OOR <	8.46	OOR <	*0.45
C9,C10	*0.25	47.58	0.96	3.87
C11,C12	OOR <	OOR <	OOR <	OOR <
D1,D2	*0.19	27.14	0.52	2.17
D3,D4	OOR <	22.98	0.6	1.95
D5,D6	*0.48	46.6	0.95	4.44
D7,D8	10.07	8.21	OOR <	5.16
D9,D10	*0.71	50.53	1.24	3.8

Starred concentrations



OOR<



Back-fit fails – concentration can't be estimated

What causes this problem?

In biomarker discovery context

- Analytes' expected concentrations uncertain
- Results in poorly tuned dose-response curves
- Pervasive problem - percent of analytes with extrapolated or out-of-range concentrations can range from 0% - >90%
- Occurs less frequently at high end of curve

How can we analyze the data and address the investigator's research hypotheses?

Simple "solution"?

Well	IL-1b	IL-2	IL-4	IL-5
B7,B8	0	23.41	0.42	1.74
B9,B10	0	51.61	1.07	3.94
B11,B12	0	5.58	0	0
C1,C2	1.04	58.38	1.18	5.59
C3,C4	0	63.17	1.66	5.69
C5,C6	0	25.23	0.38	2.32
C7,C8	0	8.46	0	0
C9,C10	0	47.58	0.96	3.87
C11,C12	0	0	0	0
D1,D2	0	27.14	0.52	2.17
D3,D4	0	22.98	0.6	1.95
D5,D6	0	46.6	0.95	4.44
D7,D8	10.07	8.21	0	5.16
D9,D10	0	50.53	1.24	3.8

Simple “solution”?

Well	IL-1b	IL-2	IL-4	IL-5
B7,B8	0	23.41	0.42	1.74
B9,B10	0	51.61	1.07	3.94
B11,B12	0	5.58	0	0
C1,C2	1.04	58.38	1.18	5.59
C3				
C5				
C7				
C9				
C1				
D1,D2	0	27.14	0.52	2.17
D3,D4	0	22.98	0.6	1.95
D5,D6	0	46.6	0.95	4.44
D7,D8	10.07	8.21	0	5.16
D9,D10	0	50.53	1.24	3.8

- Induces bias
- Underestimates variability
- Inflates type I error

Other ideas?

- Use established methods for analysis of left-censored data
  - ML with left-censored values contributing 1 - CDF to likelihood
  - Harrell's C-index - AUC equivalent for censored time-to-event data
- *Where to censor?*
- Conventional approach - censor at LOD or LOQ
- No established assay detection/quantitation limits
- Would need to be determined for each analyte and plate
- Previous efforts resulted in high rates of censored data - not an agreeable solution to the scientific investigator
- Shouldn't we account for uncertainty in curve and background estimation?

A "unified" approach

Incorporate background estimation, standard curve fitting, and patient analyte concentration estimation into a single Bayesian hierarchical model

1. *Background*

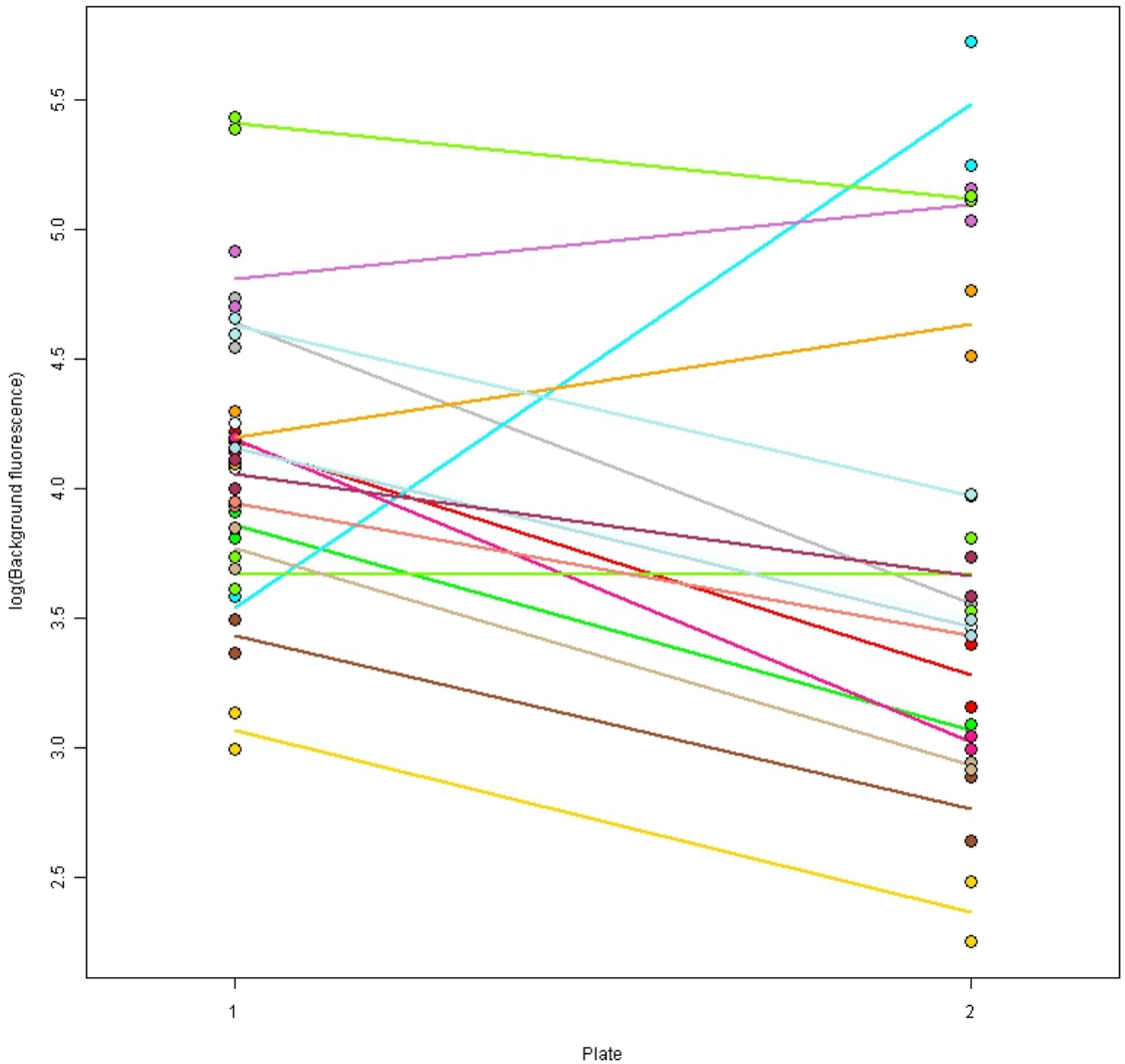
   $\delta_{jk} =$ observed background for analyte j, plate k

   $$\log(\delta_{jk}) \sim \text{Normal}(\mu_{jk}, \varphi)$$

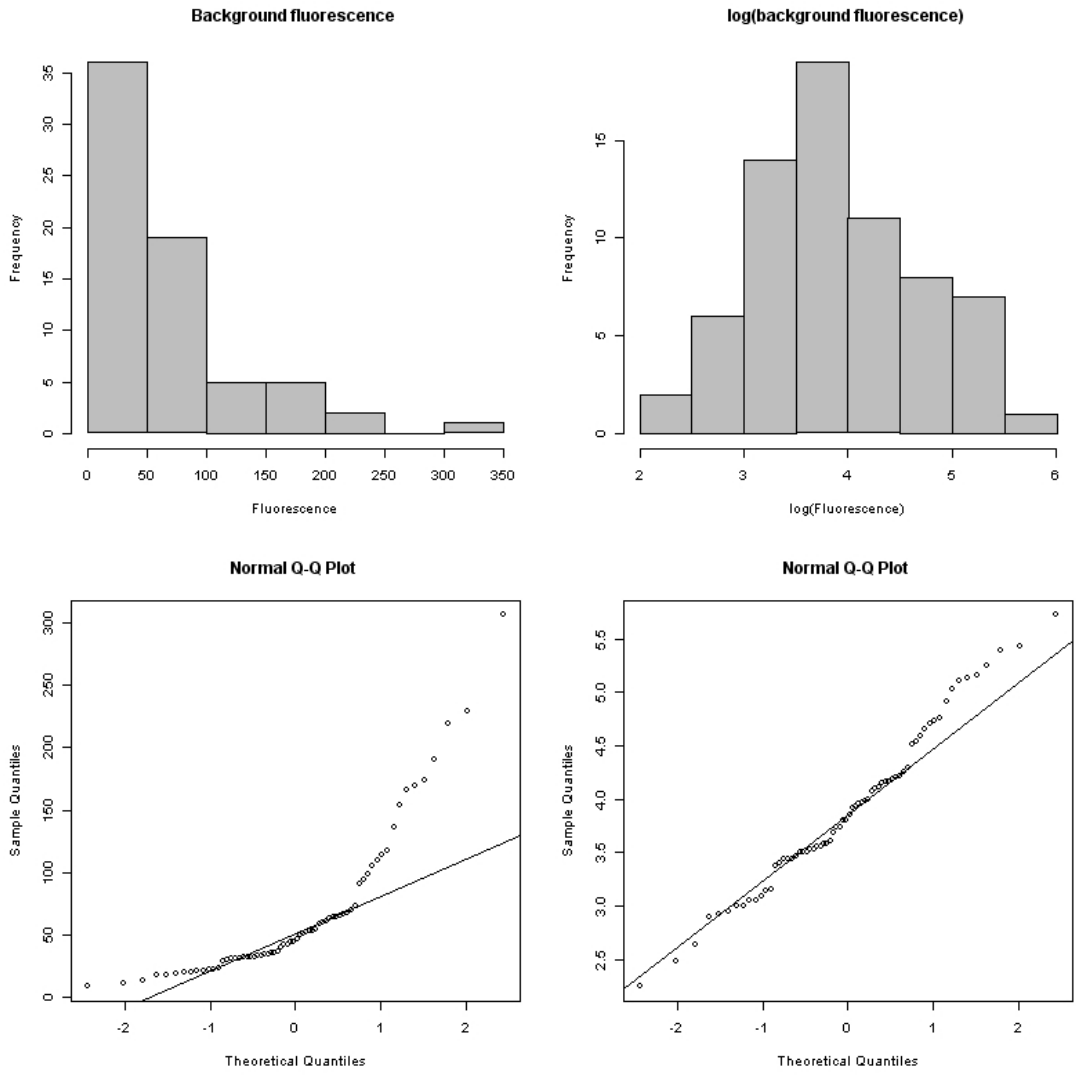
   with

   $$E(\delta_{jk}) = e^{\mu_{jk}}$$
   $$\text{Var}(\delta_{jk}) = \lambda_{jk} = e^{2\mu_{jk}} e^{\varphi}(e^{\varphi} - 1)$$

Background mean



Background distribution



Standard curve

2. *Standard curve*

$y_{ijk\ell}$ = observed (not BgC) response corresponding to the ℓth replicate of x_{ijk}, the ith known concentration for analyte j, plate k

$$y_{ijk\ell}|\mu_{jk} \sim \text{Normal}(f(x_{ijk}|\boldsymbol{\beta}_{jk}) + e^{\mu_{jk}}, \tau_{ijk})$$

where

$$\tau_{ijk} = \sigma^2_{jk} f(x_{ijk}|\boldsymbol{\beta}_{jk})^{2\theta}$$

Patient data

3. *Patient analyte concentrations*

$\nu_{hj} =$ patient h's true (latent) concentration of analyte j

$$\log(\nu_{hj}) \sim \text{Normal}(\gamma_{0j} + \gamma_{1j} D_h, \eta_j)$$

where

$$D_h = \begin{cases} 1, & \text{patient } h \text{ is disease positive} \\ 0, & \text{patient } h \text{ is disease negative.} \end{cases}$$

$e^{\gamma_{1j}} =$ fold-change comparing Ds+ to Ds- patients

Patient data

4. *Patient response values*

$z_{hjk\ell}$ = observed (not BgC) response corresponding to the ℓth replicate of patient h's true concentration for analyte j, plate k

$$z_{hjk\ell}|\mu_{jk} \sim \text{Normal}(f(\nu_{hj}|\beta_{jk}) + e^{\mu_{jk}}, \xi_{hjk})$$

where

$$\xi_{hjk} = \sigma_{jk}^2 f(\nu_{hj}|\beta_{jk})^{2\theta} \cdot 1(\nu_{hj} \geq x_{1hj}) + \lambda_{jk} \cdot 1(\nu_{hj} < x_{1hj})$$

Simulation study - experimental design

- Three plates
- Ten analytes per plate (“A” - “J”)
- Ten duplicate standards for each analyte
- Two blanks per plate
- Duplicate samples for 15 disease and 15 control subjects per plate (total sample size = 45 disease and 45 control)

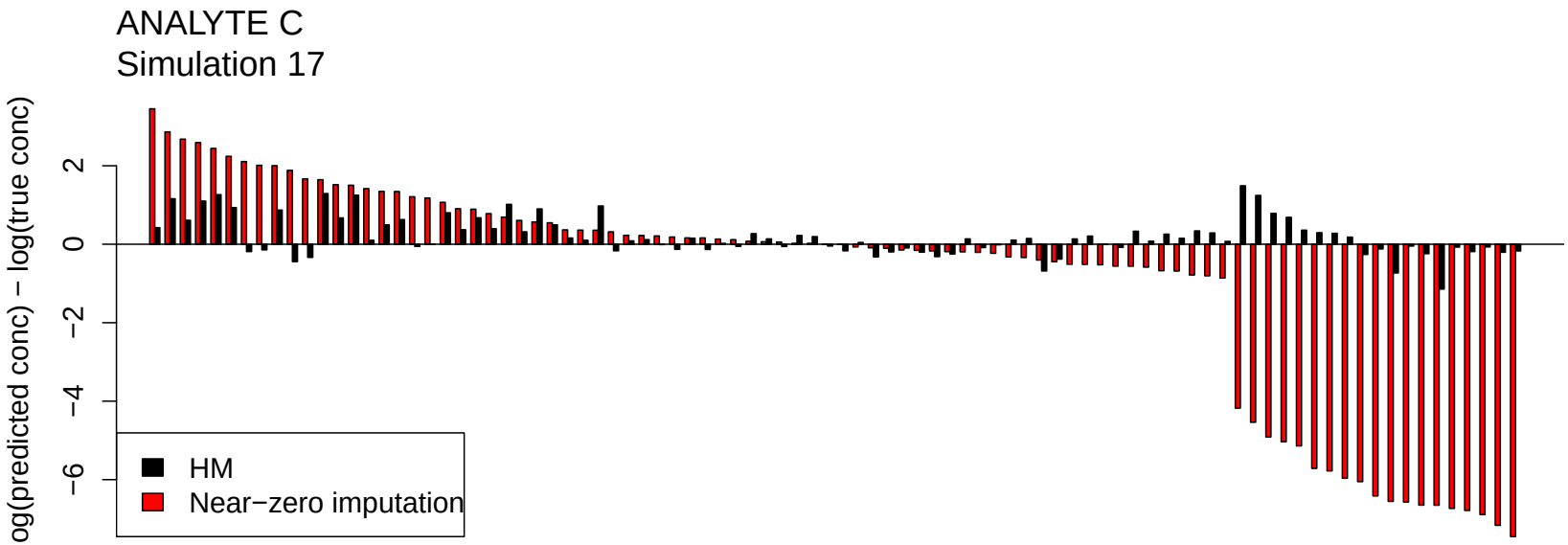
Simulation study - data description

- Prob($\nu_{hj} < x_{1jk}|$control) $= 0.2 \forall j, k$
- $e^{\gamma_{1j}} = 1, 1.5, 2, 4$, and 6, respectively, for analytes A/F, B/G, C/H, D/I, and E/J
- For control subjects, SD(log conc) = 1 for all analytes
- For disease subjects, SD(log conc) = 1 for analytes A - E
- For disease subjects, SD(log conc) = 1.25 for analytes F - J

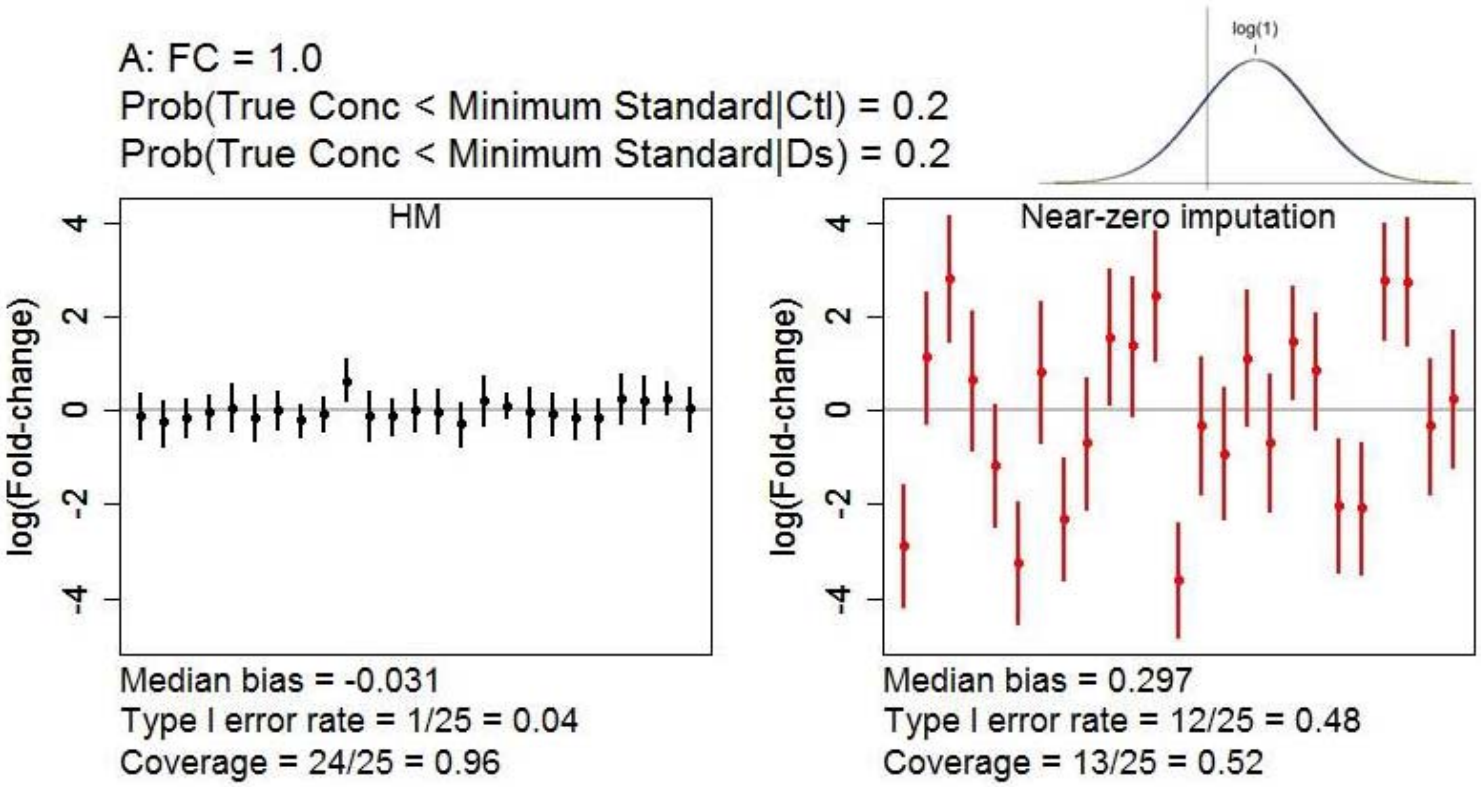
Simulation study - inference

- 25 independent simulations in WinBUGS
- Burn-in = 50K iterations with posterior inference chain length of 10K
- Constructed FC and AUC from posterior estimates of patient analyte concentrations
- Compared to standard workflow
  - "Near-zero" imputation for low extrapolated or out-of-range concentrations
  - Linear regression for FC estimation and inference
    $H_0: \text{FC} = 1$ vs $H_1: \text{FC} \neq 1$
  - Empirical estimate of AUC with inference based on corresponding 95% CI
    $H_0: \text{AUC} = 0.5$ vs $H_1: \text{AUC} > 0.5$
- Type I error, power, coverage, bias, MSE to results using standard workflow

Results

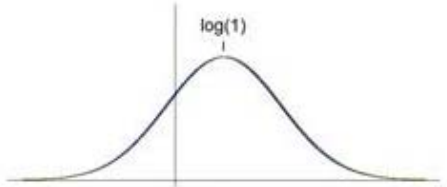

ANALYTE C
Simulation 17
log(predicted conc) − log(true conc)
2
0
−2
−4
−6
HM
Near−zero imputation

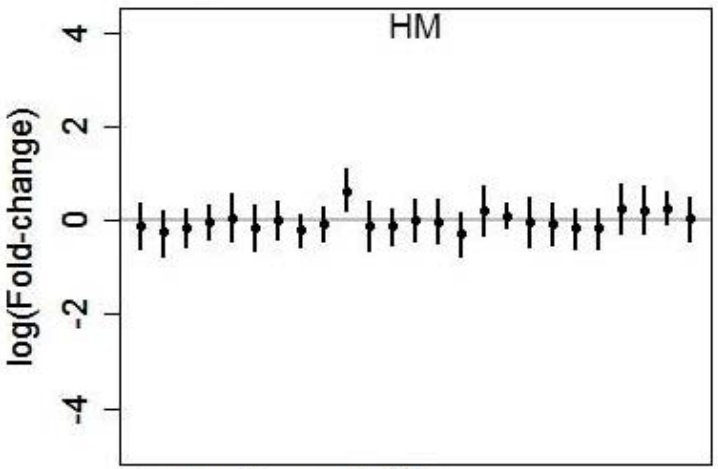
Fold change results



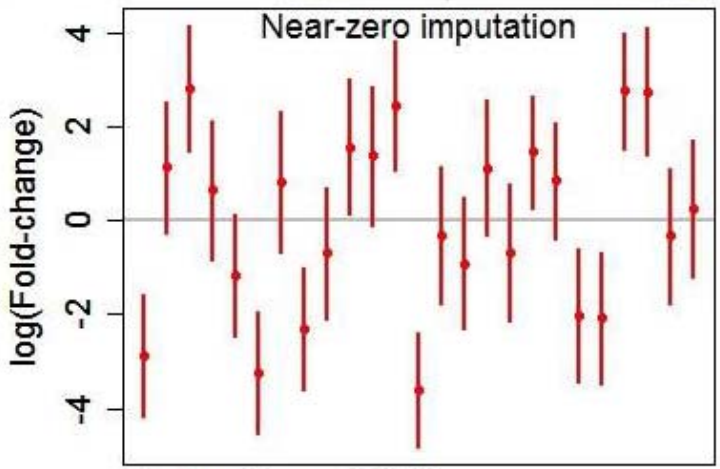
Fold change results

A: FC = 1.0
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.2





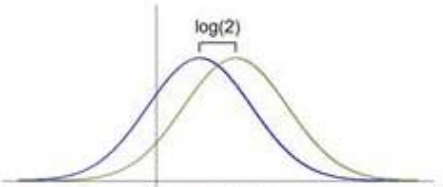
Median bias = -0.031
Type I error rate = 1/25 = 0.04
Coverage = 24/25 = 0.96

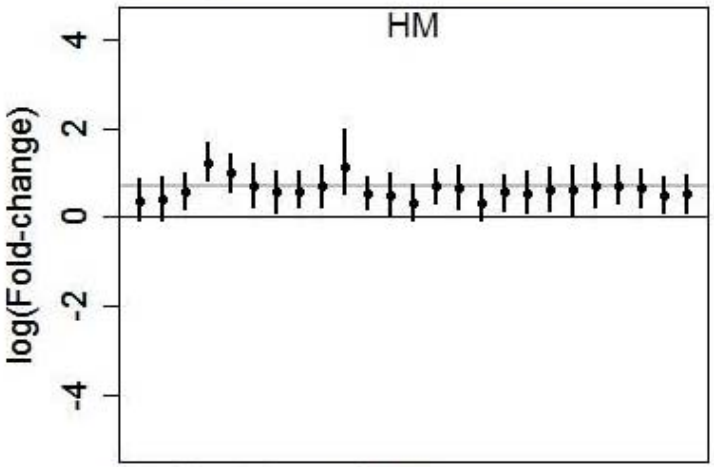


Median bias = 0.297
Type I error rate = 12/25 = 0.48
Coverage = 13/25 = 0.52

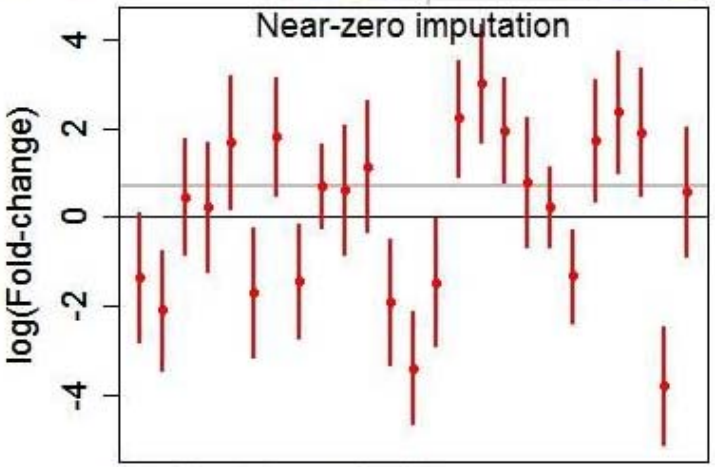
Fold change results

C: FC = 2
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.06





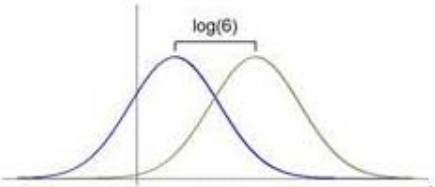
Median bias = -0.215
Power = 21/25 = 0.84
Coverage = 24/25 = 0.96

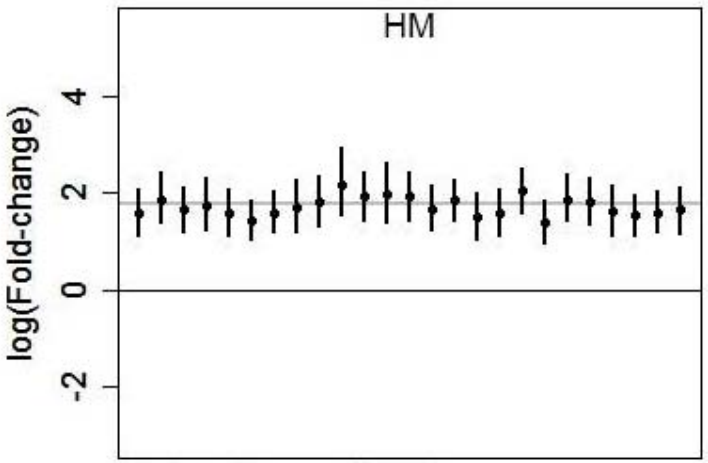


Median bias = -0.248
Power = 16/25 = 0.64
Coverage = 12/25 = 0.48

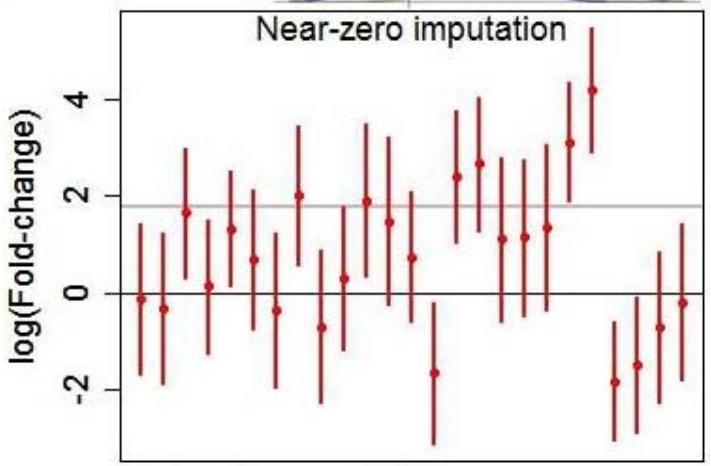
Fold change results

E: FC = 6
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.004





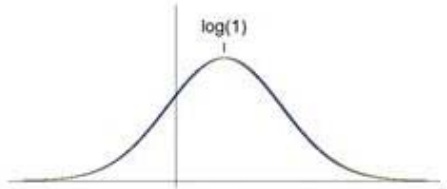
Median bias = -0.693
Power = 25/25 = 1
Coverage = 25/25 = 1

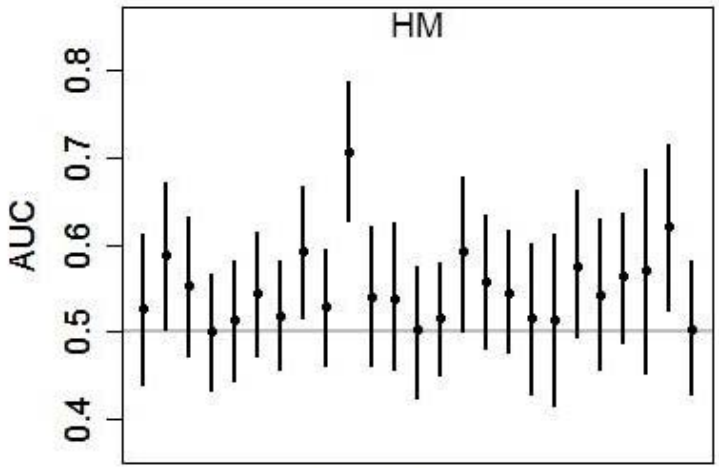


Median bias = -1.908
Power = 11/25 = 0.44
Coverage = 16/25 = 0.64

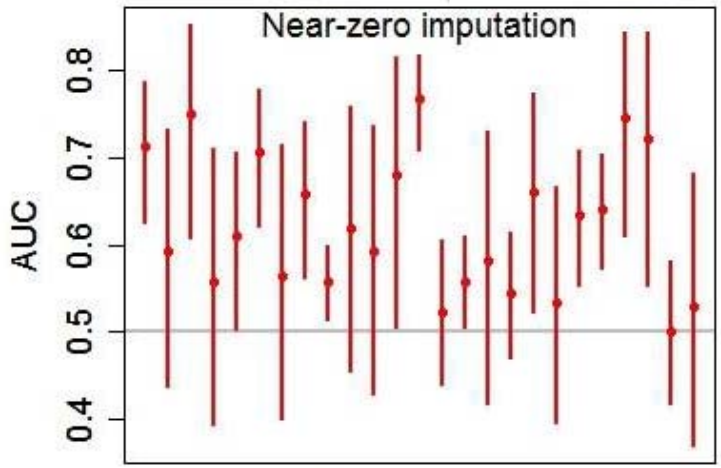
AUC results

A: AUC = 0.5
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.2



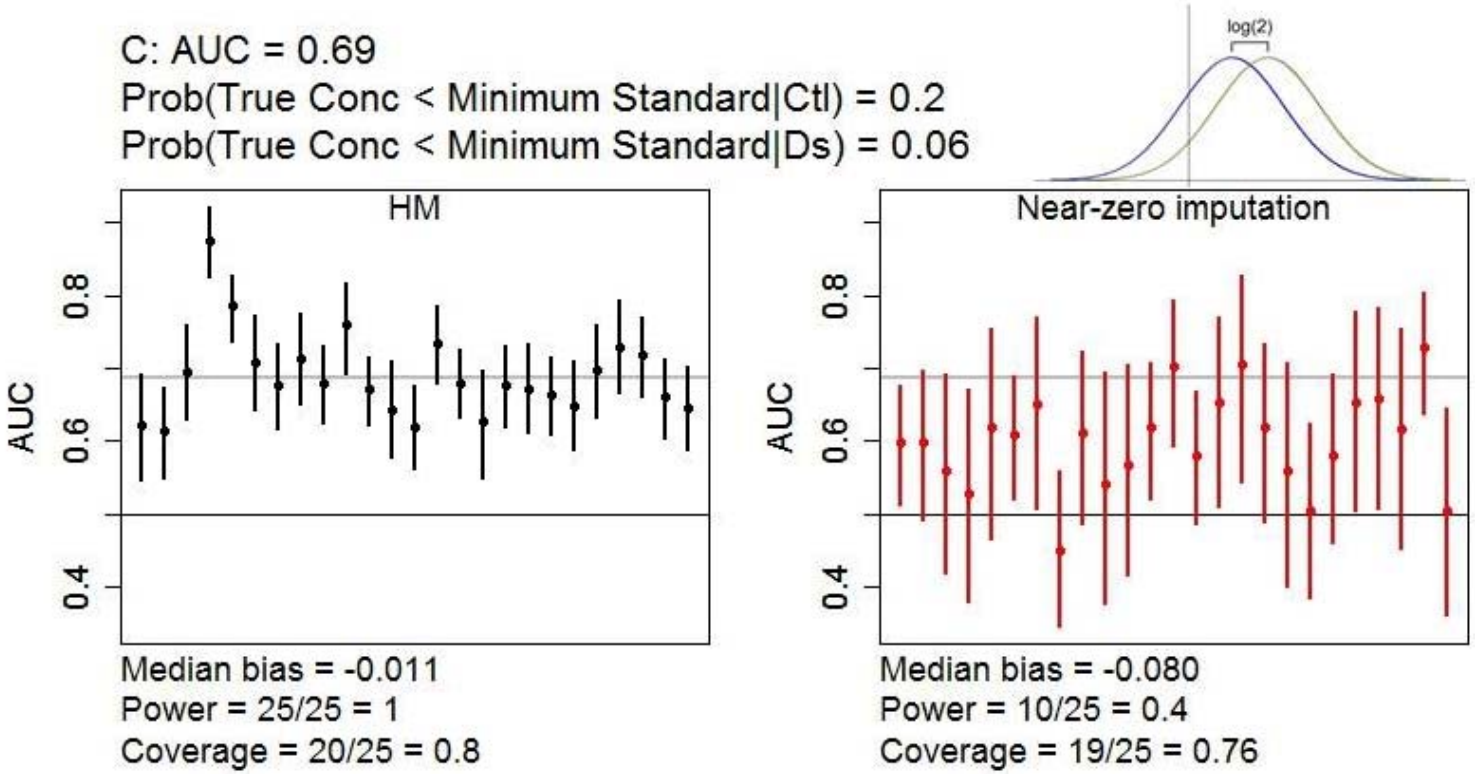


Median bias = 0.043
Type I error rate = 5/25 = 0.2
Coverage = 20/25 = 0.8

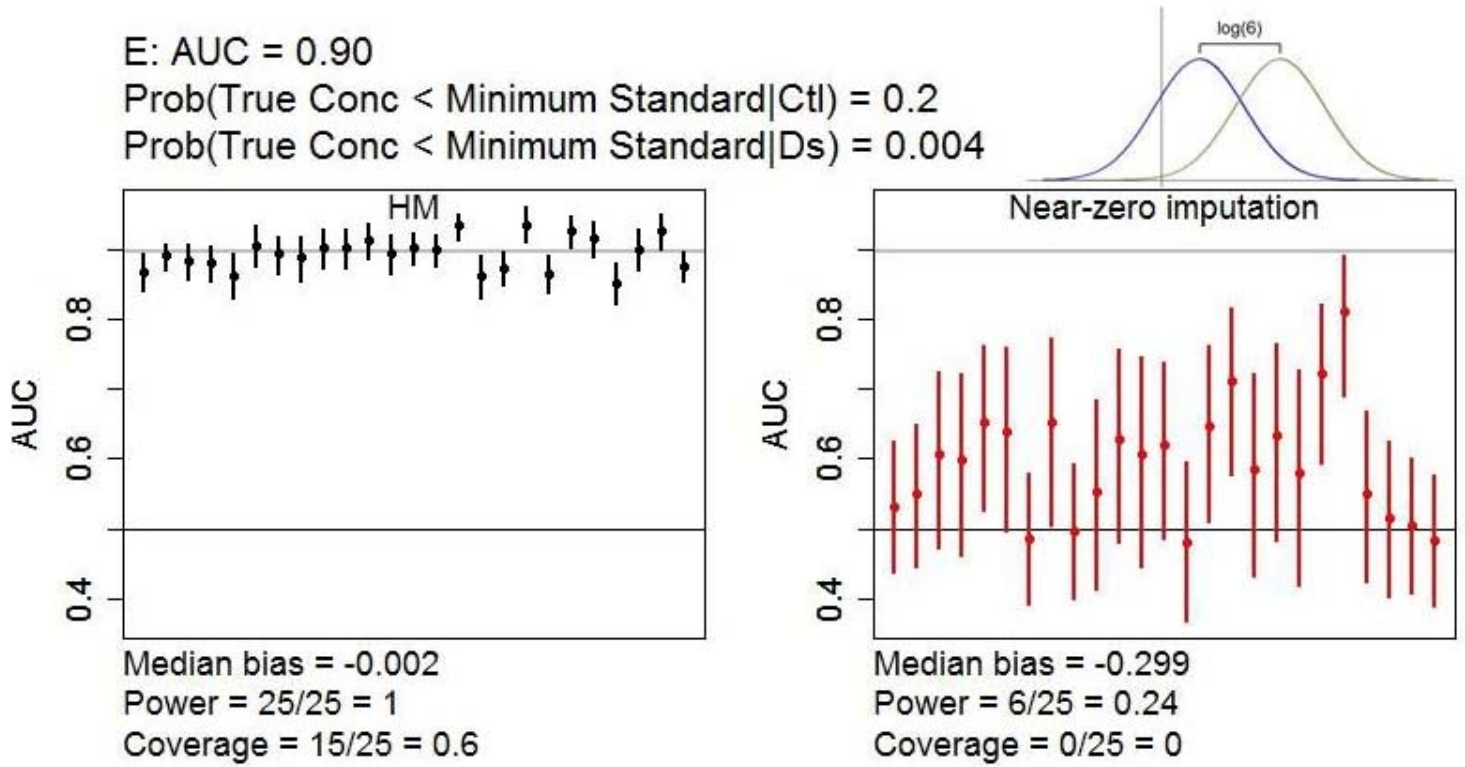


Median bias = 0.109
Type I error rate = 14/25 = 0.56
Coverage = 11/25 = 0.44

AUC results


C: AUC = 0.69
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.06
log(2)
HM
AUC
0.4
0.6
0.8
Median bias = -0.011
Power = 25/25 = 1
Coverage = 20/25 = 0.8
Near-zero imputation
AUC
0.4
0.6
0.8
Median bias = -0.080
Power = 10/25 = 0.4
Coverage = 19/25 = 0.76

AUC results


E: AUC = 0.90
Prob(True Conc < Minimum Standard|Ctl) = 0.2
Prob(True Conc < Minimum Standard|Ds) = 0.004
log(6)
HM
AUC
0.4
0.6
0.8
Median bias = -0.002
Power = 25/25 = 1
Coverage = 15/25 = 0.6
Near-zero imputation
AUC
0.4
0.6
0.8
Median bias = -0.299
Power = 6/25 = 0.24
Coverage = 0/25 = 0

Head and neck cancer application

Hypothesis: HPV16 infection affects tumor immunity, resulting in differences between patients with HPV-associated and non-associated head and neck cancer tumors.

- Serum samples from 15 HPV+, 18 HPV-, and 19 control patients
- Multiplex analysis of 9 cytokines and chemokines
- Two plates
- Eight standards per analyte and plate
- Two background wells per analyte per plate
- All patient samples measured in duplicate
- Percent of estimated concentrations below minimum standard (based on standard workflow) ranges from 0% to 41%
- FC and AUC estimated using hierarchical model

HNCa results

IL6 was 3.4-times higher in HPV- samples relative to control (FC = 3.4, 95% CI = 1.6 to 7.9).

Cytokines significantly predictive of HPV-associated and non-associated HNCa, based on AUC posterior inference. AUC (95% CI)			
Analyte	**HPV-:Ctl**	**HPV-:HPV+**	**HPV+:Ctl**
GCSF	**0.70 (0.56, 0.83)**	0.54 (0.39, 0.68)	**0.66 (0.51, 0.79)**
IL4	0.58 (0.49, 0.67)	**0.64 (0.54, 0.73)**	0.45 (0.33, 0.56)
IL6	**0.79 (0.75, 0.83)**	**0.67 (0.63, 0.71)**	**0.65 (0.59, 0.70)**
IL12	**0.61 (0.54, 0.68)**	**0.60 (0.53, 0.69)**	0.51 (0.41, 0.60)
MCP1	**0.66 (0.61, 0.70)**	**0.59 (0.54, 0.64)**	0.54 (0.48, 0.60)
MIP1b	0.52 (0.48, 0.56)	0.46 (0.42, 0.51)	**0.55 (0.51, 0.60)**
TNFa	**0.62 (0.55, 0.68)**	0.58 (0.50, 0.66)	0.55 (0.47, 0.62)

Summary

- What we've learned so far
  - If goal is biomarker discovery, near-zero imputation is a bad idea
  - FC estimation and inference has good accuracy
  - AUC estimation and inference seems slightly conservative
- What's left
  - Additional simulations
  - Compare to censored data approaches

Future directions

- Multivariate distribution for analyte concentrations
- Design considerations - number of background replicates, number of plates (sample size)
- Flexible dose-response curve fitting - e.g. splines
- Hopefully the next grant!

Acknowledgements

NIDCR/NIH R03 DE021775-01A1

Thank you!